GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 10:02:32 ; Search time 4104 Seconds
(without alignments)
16417.679 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 1647
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1647	100.0	4710	1	LIU45241	U45241 Lawsonia in
2	798.8	48.5	2672	1	XBCGROEX	M86549 Amoeba prot
3	784.8	47.7	1662	1	BAP439086	AJ439086 Buchnera
4	778.4	47.3	1811	1	LPNHTPBG	M31918 Legionella
5	777	47.2	2715	1	AB057417	AB057417 Pseudoalt
6	771.2	46.8	1635	1	BBGROELT	Z15160 B.bacillifo
7	771.2	46.8	1894	1	BAOBB63A	M98257 Bartonella
8	770.6	46.8	1644	1	BHU78514	U78514 Bartonella
9	764.6	46.4	2241	1	PHA243594	AJ243594 Pseudoalt
10	762.6	46.3	1644	1	RQU78515	U78515 Bartonella
11	756	45.9	2016	1	LPNHTPB	M91673 Legionella
12	755	45.8	1627	1	AF014829	AF014829 Bartonell
13	753	45.7	1653	1	BAP439085	AJ439085 Buchnera
14	751.2	45.6	2770	1	LMGROELSR	X57520 L.micdadei
15	746.6	45.3	1623	6	AX607355	AX607355 Sequence
16	746.6	45.3	7265	6	AX602153	AX602153 Sequence
c 17	746.6	45.3	21495	1	AE014285	AE014285 Streptoco
c 18	746.6	45.3	89050	1	SAG766855	AL766855 Streptoco
19	744	45.2	2596	1	FTGRO	X98853 F.tularensi
20	741.8	45.0	96109	6	AX067460	AX067460 Sequence
21	741	45.0	2054	1	AF389514	AF389514 Streptoco
22	739.4	44.9	1635	1	AY256822	AY256822 Piscirick
23	738.8	44.9	1647	1	BAP439083	AJ439083 Buchnera
24	738.6	44.8	1718	1	AF325222	AF325222 Streptoco
25	732.4	44.5	5907	1	D85628	D85628 Buchnera ap
26	732.4	44.5	12782	1	AE014077	AE014077 Buchnera
c 27	732.4	44.5	34699	1	AF008210	AF008210 Buchnera
28	732.2	44.5	1901	6	AX416527	AX416527 Sequence
29	729	44.3	213251	6	AX413015	AX413015 Sequence
c 30	729	44.3	258650	1	AL596171	AL596171 Listeria
c 31	729	44.3	349980	6	AX417047	AX417047 Sequence
32	727	44.1	290029	1	AE017024	AE017024 Bacillus
33	725.8	44.1	11763	1	U32736	U32736 Haemophilus
34	725.8	44.1	110000	6	AR274513_05	Continuation (6 of
35	724.4	44.0	2021	1	BSP344977	AJ344977 Buchnera
36	724.4	44.0	2028	1	BSP344978	AJ344978 Buchnera
37	723.4	43.9	300169	1	AE016801	AE016801 Vibrio vu
38	722.6	43.9	2156	1	AF335323	AF335323 Listeria
c 39	722.6	43.9	11149	1	AE015019	AE015019 Streptoco
c 40	722.6	43.9	295050	1	AL591982	AL591982 Listeria
c 41	722.6	43.9	349980	6	AX641670	AX641670 Sequence
c 42	722.6	43.9	349980	6	AX641671	AX641671 Sequence
43	721.2	43.8	1647	1	AF387863	AF387863 Buchnera
44	721.2	43.8	2200	1	HEAGROESL	M91030 Haemophilus
45	719.6	43.7	1647	1	AF434719	AF434719 Buchnera

ALIGNMENTS

RESULT 1
LIU45241
LOCUS LIU45241 4710 bp DNA linear BCT 23-SEP-1998
DEFINITION Lawsonia intracellularis plasmid pISI-2 GTP-binding protein homolog gene, partial cds, L27 50S ribosomal binding protein homolog, L21 50S ribosomal binding protein homolog, GroES/HSP10 homolog, and GroEL/HSP60 homolog genes, complete cds.
ACCESSION U45241
VERSION U45241.1 GI:3599919
KEYWORDS .
SOURCE Lawsonia intracellularis
ORGANISM Lawsonia intracellularis
Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales; Desulfovibrionaceae; Lawsonia.
REFERENCE 1 (bases 1 to 4710)

AUTHORS Dale,C.J., Moses,E.K., Ong,C.C., Morrow,C.J., Reed,M.B., Hasse,D. and Strugnell,R.A.
TITLE Identification and sequencing of the groE operon and flanking genes of Lawsonia intracellularis: use in phylogeny
JOURNAL Microbiology 144 (Pt 8), 2073-2084 (1998)
MEDLINE 98386497
PUBMED 9720028
REFERENCE 2 (bases 1 to 4710)
AUTHORS Dale,C.J.H.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-1996) C. Jane H. Dale, Molecular Biology, Victorian Institute of Animal Science, 475 Mickleham Road, Attwood, VIC 3049, Australia

FEATURES Location/Qualifiers

source 1. .4710
/organism="Lawsonia intracellularis"
/mol_type="genomic DNA"
/db_xref="taxon:29546"
/plasmid="pISI-2"

CDS complement(<1. .345)
/codon_start=1
/transl_table=11
/product="GTP-binding protein homolog"
/protein_id="AAC36496.1"
/db_xref="GI:3599920"
/translation="MRFVDEVTISVSAGKGGNGCVSFRREKFIPKGGPNGGDGGDGGN
IIFKADSRLLTLYDFRVQRHYRAQNGEGGKGSQRHGKKGEDLILHLPVGTIIFEQLLD
KEYFLVDLDRPGV"

CDS complement(508. .774)
/codon_start=1
/transl_table=11
/product="L27 50S ribosomal binding protein homolog"
/protein_id="AAC36497.1"
/db_xref="GI:3599921"
/translation="MAHKKAGGSSRNGRDSPGQRRGIKRFGGQKVLAGNILVRQLGTK
IHPGENVGMGRDYTLFAKVNGIVTYETYFRKCKRYSRVHIQPTV"

CDS complement(787. .1095)
/codon_start=1
/transl_table=11
/product="L21 50S ribosomal binding protein homolog"
/protein_id="AAC36498.1"
/db_xref="GI:3599922"
/translation="MYAIIEAGGKQFCVEEGSKIFVSKIDAEVGTEIFFDKIFMIGGS
SPQIGTPYINNAKVIAKVLEHGRDKKILVFKKWRRNDSRKLQGHRQDYTALKVTGIQL
"

CDS 1595. .1900
/note="heat shock protein; chaperonin"
/codon_start=1
/transl_table=11
/product="GroES/HSP10 homolog"
/protein_id="AAC36499.1"
/db_xref="GI:3599923"
/translation="MNLKPLNDRVLVKRLESEEKTAGGLYIPDTAKEKPSRGEVVAVG
PGKHTDDGKLIPMAVKAGDTVLFNKYAGTEVKLDGVEHLVMREDDILAVITGETGRK"

CDS 2048. .3694
/note="heat shock protein; chaperonin"
/codon_start=1
/transl_table=11
/product="GroEL/HSP60 homolog"
/protein_id="AAC36500.1"
/db_xref="GI:3599924"
/translation="MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSF
GSPVITKDGVSVAKEIELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYRE
GVKLVAAGRNPMAIKRGIDKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGN
IIAEAMAKVGKGGVITVEEAKGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNP
YILCNEKKITSMKDMLPILEQVAKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAV
KAPGFGERRKAMLEDIAILTGGEAIFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVD
GAGKSEDIKARVKQIRAQIEETSSDYDREKLQERLAKLVGGVAVIHVGAATETEMKEK
KDRVEDALNATRAAVEEGIVPGGGTAFVRSIKVLDDIKPADDDELAGLNIIRRSLEEP
LRQIAANAGYEGSIVVEKVREPKDGFGFNAASGEYEDLIKAGVIDPKKVTRIALQNAA
SVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMGGMDGMY"

CDS 3821. .4702
/note="orfx; similar to Escherichia coli possible protoporphyrinogen oxidase (hemK)"
/codon_start=1
/transl_table=11
/protein_id="AAC36501.1"
/db_xref="GI:3599925"
/translation="MIKYIDWIQKTTMMFKKAGIDSPKLSAELILSHVLNITRLQIIM
TPFEPIPTNSYSTLNDIMLRRLHGEPIAYLTGKKEFFSREFKVTQATLIPRPETELLI
EFVLNHINPTQQIYFADLGTGSGCIAITLAAERKNWLGIATDISSEALKIAKLNSLKN
NTHSQLQFLQSDFTQPLCLPSSLDLYISNPPYISENELTSLPHEVISFEPKIALTPHK
CIHLDEINTVLHCYKKIITQAEISLKPGGIIILEHGATQAEAILLLLKNNIWTNVISH
TDLTNKNRFITAYKYKI"

BASE COUNT 1638 a 827 c 785 g 1460 t
ORIGIN

Query Match 100.0%; Score 1647; DB 1; Length 4710;
Best Local Similarity 100.0%; Pred. No. 2.4e-283;
Matches 1647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2048 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 2107

Qy      61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2108 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 2167

Qy     121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2168 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 2227

Qy     181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2228 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 2287

Qy     241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2288 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 2347

Qy     301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2348 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 2407

Qy     361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2408 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 2467

Qy     421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2468 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 2527

Qy     481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2528 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 2587

Qy     541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2588 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 2647

Qy     601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2648 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 2707

Qy     661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2708 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 2767

Qy     721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2768 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 2827

Qy     781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2828 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 2887
```

```
Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2888 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 2947

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2948 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 3007

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3008 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 3067

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3068 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 3127

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3128 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 3187

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3188 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 3247

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3248 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 3307

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3308 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 3367

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3368 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 3427

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3428 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 3487

Qy      1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3488 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 3547

Qy      1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3548 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 3607

Qy      1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3608 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 3667

Qy      1621 ATGGGTGGTATGGACGGTATGTACTAG 1647
            |||||||||||||||||||||||||||
Db      3668 ATGGGTGGTATGGACGGTATGTACTAG 3694
```

```
RESULT 2
XBCGROEX
LOCUS       XBCGROEX                2672 bp    DNA     linear   BCT 02-APR-1998
DEFINITION  Amoeba proteus symbiotic bacterium groES and groEL genes, complete
            cds's.
ACCESSION   M86549
VERSION     M86549.1  GI:2996603
KEYWORDS    .
SOURCE      Amoeba proteus symbiotic bacterium
  ORGANISM  Amoeba proteus symbiotic bacterium
            Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
            Legionellaceae.
REFERENCE   1  (bases 1 to 2672)
  AUTHORS   Ahn,T.I., Leeu,H.K., Kwak,I.H. and Jeon,K.W.
  TITLE     Nucleotide sequence and temperature-dependent expression of XgroEL
            gene isolated from symbiotic bacteria of Amoeba proteus
  JOURNAL   Endocyt. Cell Res. 8, 33-44 (1991)
REFERENCE   2  (sites)
  AUTHORS   Ahn,T.I., Lim,S.T., Leeu,H.K., Lee,J.E. and Jeon,K.W.
  TITLE     A novel strong promoter of the groEx operon of symbiotic bacteria
            in Amoeba proteus
  JOURNAL   Gene 148 (1), 43-49 (1994)
  MEDLINE   95011657
   PUBMED   7926836
REFERENCE   3  (bases 1 to 2672)
  AUTHORS   Jeon,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1992) Kwang W. Jeon, Zoology, University of
            Tennessee, Cumberland Street, Knoxville, TN 37996-0810
COMMENT     On Mar 30, 1998 this sequence version replaced gi:155400.
FEATURES             Location/Qualifiers
     source          1. .2672
                     /organism="Amoeba proteus symbiotic bacterium"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:2728"
     gene            232. .691
                     /gene="groES"
     promoter        232. .262
                     /gene="groES"
                     /note="P1"
                     /citation=[1]
     -35_signal      272. .279
                     /gene="groES"
                     /note="-35-P2a"
                     /citation=[1]
     -10_signal      294. .302
                     /gene="groES"
                     /note="-10-P1"
                     /citation=[1]
     RBS             387. .390
                     /gene="groES"
                     /note="SD"
                     /citation=[1]
     CDS             401. .691
                     /gene="groES"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC09380.1"
                     /db_xref="GI:155401"
                     /translation="MKIRPLHDRVVVRRLEEERTTAGWIVIPDSATEKPMRGEIIAIG
                     AGKILDNGDVRAFVVKVGDVVLFGKYSGTEVKVAGQELVVMREDDIMGVIEK"
     gene            566. .2373
                     /gene="groEL"
     -10_signal      566. .572
                     /gene="groES"
                     /note="-10-P2a"
                     /citation=[1]
     promoter        582. .631
                     /gene="groES"
     -35_signal      582. .588
                     /gene="groES"
                     /citation=[1]
     RBS             704. .707
                     /gene="groEL"
                     /note="SD"
                     /citation=[1]
     CDS             718. .2373
                     /gene="groEL"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC09381.1"
                     /db_xref="GI:155402"
                     /translation="MAKELRFGDDARQQMLAGVNALADRVKATMGPSGRNVVLERSFG
                     APTVTKDGVSVAKEIEFENRFKNMGAQMVKEVAAKTSDTAGDGTTTATVLARSIVVEG
                     HKAVAAGMNPMDLKRGIDKAVTAITKELQKMSKPCKDGKAIAQVGTISANSDQAIGSI
                     IAEAMEKVGKEGVITVEDGNGLENELSVVEGMQFDRGYISPYFINNQQNMSAELEHPF
                     ILLVDKKIATIRDMLSVLEAVAKSGRPLLIVAEDVEGEALATLVVNNMRGIVKVCAVK
                     APGFGDRRKAMLQDIAILTNGQVISEEIGTSLETASLESLGTAKRIVVTKENTTIIDG
```

```
                    EGKATEINARIAQIRAQMEETSSDYDREKLQERVAKLAGGVAVIKVGAATEIEMKEKK
                    ARVEDALHATRAAVEEGIVAGGGVALIRAQKVLDGLKGDNADQDMGINILRRAIESPL
                    RQIVANAGYESSVIVNKVAEHKDNFGFNAATGQYGDMVEMGILDPTKVTRTALQNAAS
                    VRSLMLTTECMVADLPKKDEGMAGAGDMGGMGGMGGMGGMGMM"
     stem_loop      2313...2332
                    /gene="groEL"
                    /citation=[1]
     stem_loop      2398..2425
                    /citation=[1]
BASE COUNT      795 a     525 c     628 g     724 t
ORIGIN

  Query Match             48.5%;  Score 798.8;  DB 1;  Length 2672;
  Best Local Similarity   68.5%;  Pred. No. 2.4e-132;
  Matches 1119;  Conservative    0;  Mismatches  512;  Indels    3;  Gaps    1;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db        722 CTAAAGAATTACGTTTTGGTGACGACGCTCGCCAACAAATGCTTGCTGGTGTGAATGCAT 781

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db        782 TGGCTGATCGGGTTAAAGCAACTATGGGCCCAAGTGGACGTAACGTTGTATTAGAAAGAT 841

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        842 CTTTTGGTGCTCCTACTGTTACCAAAGACGGTGTTTCTGTTGCAAAAGAAATTGAATTTG 901

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        902 AAAACCGTTTCAAAAACATGGGCGCACAAATGGTTAAAGAAGTTGCTGCTAAAACTTCTG 961

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        962 ACACCGCTGGTGACGGTACAACAACTGCGACTGTCTTAGCTCGCTCAATCGTTGTTGAAG 1021

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db       1022 GTCATAAAGCAGTTGCTGCTGGCATGAACCCAATGGATCTGAAACGCGGTATCGACAAAG 1081

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db       1082 CAGTTACAGCAATTACTAAAGAACTGCAAAAAATGTCCAAGCCTTGCAAAGATGGCAAAG 1141

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db       1142 CTATTGCTCAGGTTGGTACTATTTCTGCCAACTCTGACCAAGCGATTGGTTCCATCATTG 1201

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db       1202 CTGAAGCAATGGAAAAAGTAGGTAAAGAAGGCGTAATTACTGTTGAAGACGGTAATGGTC 1261

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db       1262 TGGAAAATGAATTATCTGTTGTTGAAGGTATGCAGTTTGACCGTGGTTACATTTCTCCAT 1321

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db       1322 ACTTCATCAACAACCAACAAAACATGAGTGCTGAGCTTGAGCATCCATTTATCCTCTTAG 1381

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db       1382 TTGACAAGAAGATTGCTACTATCCGTGATATGTTGTCTGTTCTTGAAGCAGTTGCCAAAT 1441

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db       1442 CAGGCCGTCCTTTACTGATCGTTGCTGAAGACGTTGAAGGTGAAGCACTGGCTACTCTGG 1501

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db       1502 TTGTTAACAACATGCGTGGAATCGTTAAAGTATGTGCTGTTAAAGCACCTGGTTTTGGCG 1561

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db       1562 ATCGCCGTAAAGCAATGTTGCAAGATATTGCTATCCTGACTAACGGCCAAGTGATTTCTG 1621

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db       1622 AAGAAATTGGCACAAGCTTAGAAACCGCTTCTCTGGAAAGCTTAGGAACTGCAAAACGTA 1681

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db       1682 TCGTTGTTACCAAAGAAAACACAACCATTATCGATGGCGAAGGCAAAGCAACTGAAATCA 1741

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1742 ATGCTCGCATCGCTCAAATCCGCGCTCAAATGGAAGAAACTTCTTCTGATTATGACCGTG 1801

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1802 AGAAGTTACAAGAGCGTGTTGCCAAACTGGCTGGTGGTGTTGCGGTTATCAAAGTTGGTG 1861

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1862 CTGCTACTGAAATCGAAATGAAAGAGAAGAAAGCTCGTGTTGAAGATGCTCTGCATGCAA 1921

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1922 CTCGCGCTGCGGTTGAAGAAGGTATCGTTGCAGGTGGTGGTGTTGCCCTGATCCGTGCTC 1981

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1982 AAAAAGTATTGGATGGCCTGAAAGGTGACAATGCTGATCAAGACATGGGTATCAACATCC 2041

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       2042 TCCGTCGTGCAATCGAATCTCCTCTGCGTCAGATCGTTGCTAATGCTGGCTACGAATCAT 2101

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       2102 CTGTTATCGTTAACAAAGTTGCTGAACACAAAGATAACTTTGGTTTCAACGCTGCAACTG 2161

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       2162 GCCAATACGGTGACATGGTTGAGATGGGTATTCTTGATCCAACTAAGGTTACTCGTACTG 2221

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       2222 CCTTACAAAACGCAGCTTCTGTTCGAAGCCTGATGTTGACCACTGAGTGCATGGTTGCTG 2281

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627
Db       2282 ACCTACCTAAGAAAGATGAAGGTATGGCT---GGTGCCGGCGACATGGGCGGCATGGGTG 2338

Qy       1628 GTATGGACGGTATG 1641
Db       2339 GTATGGGCGGTATG 2352

RESULT 3
BAP439086
LOCUS       BAP439086               1662 bp    DNA     linear   BCT 12-JUN-2002
DEFINITION  Buchnera aphidicola (Tuberolachnus salignus) groEL gene.
ACCESSION   AJ439086
VERSION     AJ439086.1  GI:21425628
KEYWORDS    groEL gene; GroEL protein.
SOURCE      Buchnera aphidicola (Tuberolachnus salignus)
  ORGANISM  Buchnera aphidicola (Tuberolachnus salignus)
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Buchnera.
REFERENCE   1
  AUTHORS   Fares,M.A., Barrio,E., Sabater-Munoz,B. and Moya,A.
  TITLE     The evolution of the heat-shock protein GroEL from Buchnera, the
            primary endosymbiont of aphids, is governed by positive selection
  JOURNAL   Mol. Biol. Evol. (2002) In press
REFERENCE   2  (bases 1 to 1662)
```

```
  AUTHORS   Sabater-Munoz,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2002) Sabater-Munoz B., Insto. Cavanilles Biodiv
            & Biol Evol, Universidad de Valencia, P.O. Box 2085, Valencia
            E-46071, SPAIN
FEATURES             Location/Qualifiers
     source          1. .1662
                     /organism="Buchnera aphidicola (Tuberolachnus salignus)"
                     /mol_type="genomic DNA"
                     /specific_host="Tuberolachnus salignus aphid"
                     /db_xref="taxon:98804"
                     /country="Spain:Valencia"
     gene            1. .1662
                     /gene="groEL"
     CDS             1. .1662
                     /gene="groEL"
                     /function="molecular chaperonin"
                     /codon_start=1
                     /transl_table=11
                     /product="GroEL protein"
                     /protein_id="CAD27798.1"
                     /db_xref="GI:21425629"
                     /translation="MAAKEVKFGNDARAKMLRGVNVLADAVKVTLGPKGRNVVLDKSF
                     GAPTITKDGVSVAREIELEDKFENMGTQMVKEVASKANDAAGDGTTTATVLAQSIITE
                     GLKAVAAGMNPMDLKRGIDKAVIAAVEELRKLSVPCSDSKAIAQVGTISANSDLTVGK
                     LIAEAMGKVGKEGVITVEEGTGLQDELDVVEGMQFDRGYLSPYFINKPESGAVELENP
                     FILLADKKISNIRELLPILEAVAKAGKSLLIIAEDVEGEALATLIVNTMRGIVKVAAV
                     KAPGFGDRRKAMLQDIATLTGGTVISEEIGLELEKTILDDLGQAKRVVINKDTTIIID
                     GIGNEINIKGRIAQIHQQIEEATSDYVKEKLQERVAKLADGVAVIKVGAATEVELKEK
                     KGRVEDALHATRAAVEEGVVAGGGVALIRVAGKISNLTGDNEDQNVGIKVALRAMESP
                     LRQIVINAGEEASVIANNVKACEGSYGYNAYTEEYGDMISMGILDPTKVTRSALQYAA
                     SVAGLMITTECMITDLPKGETPDLNNSGGMGGGMGGGMGGMGGMM"
BASE COUNT      573 a    202 c    371 g    516 t
ORIGIN

  Query Match             47.7%;  Score 784.8;  DB 1;  Length 1662;
  Best Local Similarity   67.8%;  Pred. No. 8e-130;
  Matches 1098;  Conservative    0;  Mismatches  522;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 ATGGCTGCGAAAGAAGTAAAATTTGGAAATGATGCTCGTGCAAAAATGCTTCGTGGTGTA 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         61 AATGTTCTTGCTGATGCAGTAAAAGTAACTTTAGGTCCTAAAGGACGTAATGTAGTATTA 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        121 GATAAATCTTTTGGTGCTCCTACTATAACTAAAGATGGTGTATCTGTTGCTCGTGAAATT 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        181 GAATTAGAAGATAAATTTGAAAACATGGGTACACAAATGGTTAAAGAAGTAGCATCTAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        241 GCTAATGATGCAGCAGGTGATGGTACAACTACTGCAACTGTTTTAGCACAATCAATTATT 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        301 ACTGAAGGTTTAAAAGCTGTAGCTGCTGGTATGAATCCAATGGATTTAAAACGTGGTATC 360

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        361 GATAAAGCTGTAATTGCAGCAGTTGAAGAATTAAGAAAACTTTCAGTACCTTGTTCTGAT 420

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        421 TCTAAAGCTATTGCTCAAGTTGGTACTATTTCTGCAAATTCAGATTTAACTGTTGGTAAA 480

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        481 TTAATTGCAGAAGCAATGGGTAAAGTAGGTAAAGAAGGTGTAATTACTGTTGAAGAAGGT 540

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        541 ACTGGACTACAAGATGAACTTGATGTAGTAGAAGGTATGCAATTTGATCGTGGTTATCTC 600

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        601 TCTCCTTACTTTATAAATAAACCAGAATCTGGTGCTGTAGAATTAGAAAATCCATTTATT 660

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        661 TTGTTAGCAGATAAAAAAATATCAAATATCCGTGAATTATTACCAATTTTAGAAGCTGTA 720

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        721 GCTAAAGCTGGAAAATCATTACTGATTATTGCTGAAGATGTAGAAGGTGAAGCATTAGCA 780

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        781 ACTTTAATAGTTAATACAATGCGTGGTATCGTAAAAGTTGCTGCAGTTAAAGCTCCTGGT 840

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        841 TTTGGAGATCGTAGAAAAGCTATGTTGCAAGATATCGCAACTTTAACTGGTGGTACAGTA 900

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        901 ATTTCTGAAGAAATTGGTTTAGAATTAGAAAAAACAATTTTAGATGATTTAGGTCAAGCA 960

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        961 AAACGTGTAGTAATTAATAAAGATACTACAATTATTATTGATGGTATTGGTAATGAAATC 1020

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1021 AATATTAAAGGACGAATTGCTCAAATACATCAACAAATTGAAGAAGCAACTTCTGATTAT 1080

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1081 GTTAAAGAAAAATTACAAGAACGTGTAGCAAAACTAGCTGATGGAGTTGCTGTTATAAAA 1140

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1141 GTTGGTGCAGCTACTGAAGTAGAATTGAAAGAAAAAAAAGGTCGTGTTGAAGATGCTTTA 1200

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1201 CATGCAACTCGTGCTGCTGTAGAAGAAGGTGTTGTAGCTGGTGGTGGTGTTGCATTAATT 1260

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1261 CGCGTAGCAGGAAAAATTAGTAATTTAACAGGCGATAATGAAGATCAAAATGTAGGTATT 1320

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1321 AAAGTTGCATTACGTGCAATGGAATCTCCACTTCGTCAAATTGTTATTAATGCTGGTGAA 1380

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       1381 GAAGCTTCTGTAATTGCTAATAATGTAAAAGCATGTGAAGGTAGTTATGGTTATAATGCT 1440

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       1441 TATACTGAAGAATATGGTGATATGATTTCAATGGGTATATTAGATCCAACTAAAGTCACC 1500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       1501 CGTTCAGCTTTACAATATGCAGCTTCTGTAGCTGGTTTAATGATTACTACTGAATGTATG 1560

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db       1561 ATCACTGATTTACCAAAAGGTGAAACACCAGATTTAAATAATTCTGGTGGTATGGGCGGT 1620
```

```
RESULT 4
LPNHTPBG
LOCUS       LPNHTPBG              1811 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Legionella pneumonphila htp heat shock operon HtpB (hptB gene),
            complete cds.
ACCESSION   M31918
VERSION     M31918.1  GI:149691
KEYWORDS    htp heat shock operon.
SOURCE      Legionella pneumophila
  ORGANISM  Legionella pneumophila
            Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
            Legionellaceae; Legionella.
REFERENCE   1  (bases 1 to 1811)
  AUTHORS   Hoffman,P.S., Houston,L. and Butler,C.A.
  TITLE     Legionella pneumophila htpAB heat shock operon: nucleotide sequence
            and expression of the 60-kilodalton antigen in L.
            pneumophila-infected HeLa cells
  JOURNAL   Infect. Immun. 58 (10), 3380-3387 (1990)
  MEDLINE   90382960
   PUBMED   2205580
COMMENT     Original source text: Legionella pneumophila (strain SVir)
            (library: cosmid clone pHC79) DNA.
            Draft entry and computer-readable sequence for [Unpublished (1990)
            Univ. of TN at Memphis, Memphis TN 38163] kindly submitted by P.S.
            Hoffman, 05-FEB-1990.
FEATURES             Location/Qualifiers
     source          1..1811
                     /organism="Legionella pneumophila"
                     /mol_type="genomic DNA"
                     /strain="SVir"
                     /db_xref="taxon:446"
                     /tissue_lib="cosmid clone pHC79"
     gene            34..1680
                     /gene="htpB"
     CDS             34..1680
                     /gene="htpB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA25299.1"
                     /db_xref="GI:149692"
                     /translation="MAKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYG
                     APTVTKDGVSVAKEIEFEHRFMNMGAQMVKEVASKTSDTAGDGTTATVLARSILVEG
                     HKAVAAGMNPMDLKRGIDKAVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAI
                     IAEAMEKVGKEGVITVEDGNGLENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPF
                     ILLVDKKVSSIREMLSVLEGVAKSGRPLLIIAEDVEGEALATLVVNNMRGIVKVCAVK
                     APGFGDRRKAMLQDIAILTKGQVISEEIGKSLEGATLEDLGSAKRIVVTKENTTIIDG
                     EGKATEINARIAQIRAQMEETTSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKK
                     ARVEDALHATRAAVEEGIVAGGGVALIRAQKALDSLKGDNDDQNMGINILRRAIESPM
                     RQIVTNAGYEASVVVNKVAEHKDNYGFNAATGEYGDMVEMGILDPTKVTRMALQNAAS
                     VASLMLTTECMVADLPKKEEGVGAGDMGGMGGMGGMGGMM"
     repeat_unit     1699..1780
                     /rpt_type=inverted
BASE COUNT      531 a    299 c    448 g    533 t
ORIGIN

  Query Match             47.3%;  Score 778.4;  DB 1;  Length 1811;
  Best Local Similarity   67.7%;  Pred. No. 1.1e-128;
  Matches 1107;  Conservative    0;  Mismatches  521;  Indels    6;  Gaps    1;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db         38 CTAAAGAATTACGTTTTGGTGATGACGCTCGCCTACAAATGCTTGCTGGTGTTAATGCAT 97

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db         98 TAGCAGATGCGGTTCAAGTTACTATGGGTCCACGTGGCCGTAATGTTGTATTGGAAAAAT 157

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        158 CTTATGGCGCTCCTACTGTAACTAAAGACGGTGTGTCTGTTGCCAAAGAAATTGAATTTG 217

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        218 AGCATCGTTTCATGAACATGGGCGCTCAAATGGTTAAAGAAGTGGCTTCTAAAACTTCTG 277

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        278 ATACTGCTGGTGATGGTACTACTACTGCAACAGTATTGGCTCGTTCTATTCTTGTTGAAG 337

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        338 GTCACAAAGCAGTTGCTGCTGGTATGAATCCAATGGATCTCAAACGCGGTATTGATAAAG 397

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        398 CAGTATTAGCAGTTACCAAAAAATTACAAGCTATGTCTAAGCCATGCAAAGACAGCAAAG 457

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        458 CCATTGCTCAAGTTGGAACTATTTCTGCTAATTCCGATGAAGCGATTGGTGCTATCATTG 517

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        518 CTGAAGCAATGGAAAAAGTTGGTAAAGAGGGTGTTATTACCGTTGAAGATGGTAATGGAT 577

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db        578 TGGAAAATGAGCTTTATGTTGTTGAAGGTATGCAATTTGATCGCGGTTACATTTCTCCAT 637

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db        638 ACTTCATCAACAACCAGCAAAACATGAGCTGTGAACTTGAGCATCCATTCATTTTATTGG 697

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db        698 TTGACAAAAAAGTTTCCAGTATTCGTGAAATGTTGTCCGTATTGGAAGGTGTTGCTAAAT 757

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db        758 CTGGCCGTCCTTTGTTGATCATCGCAGAAGATGTTGAAGGCGAAGCTTTAGCTACTCTGG 817

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db        818 TAGTCAACAACATGCGCGGTATTGTAAAAGTATGTGCTGTCAAAGCGCCTGGTTTTGGTG 877

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db        878 ATCGTCGCAAAGCGATGTTGCAAGACATTGCTATTTTGACTAAGGGTCAAGTTATTTCTG 937

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db        938 AAGAAATTGGCAAGAGCCTGGAAGGTGCTACTCTGGAAGATCTTGGTAGTGCTAAGCGAA 997

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db        998 TCGTTGTTACCAAAGAAAACACTACTATCATTGATGGTGAAGGAAAGGCAACTGAAATTA 1057

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1058 ATGCTCGTATTGCTCAAATTCGTGCACAAATGGAAGAAACCACTTCTGATTACGATAGAG 1117

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1118 AAAAATTACAAGAGCGCGTTGCTAAGTTAGCTGGTGGTGTTGCTGTTATCAAAGTTGGCG 1177

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1178 CTGCCACAGAAGTTGAAATGAAAGAGAAGAAAGCTCGTGTTGAAGATGCTCTTCATGCTA 1237

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1238 CTCGCGCTGCAGTAGAAGAAGGTATCGTTGCCGGTGGTGGCGTTGCCTTGATTCGTGCTC 1297

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
```

```
Db       1298 AGAAAGCTCTTGATTCATTGAAAGGTGATAATGACGATCAAAATATGGGTATCAATATTT 1357

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387

Db       1358 TACGTCGCGCTATTGAATCTCCAATGCGTCAGATTGTTACTAACGCAGGATATGAAGCTT 1417

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447

Db       1418 CTGTTGTAGTAAACAAGGTAGCTGAGCACAAAGACAACTACGGTTTCAACGCTGCAACTG 1477

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507

Db       1478 GTGAATACGGTGATATGGTTGAGATGGGTATTCTTGATCCAACTAAAGTAACCCGTATGG 1537

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567

Db       1538 CTCTGCAAAATGCAGCTTCTGTAGCCAGTTTGATGTTGACTACTGAATGTATGGTTGCTG 1597

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627

Db       1598 A------TCTGCCTAAGAAAGAAGAAGGTGTTGGTGCCGGCGATATGGGCGGCATGGGCG 1651

Qy       1628 GTATGGACGGTATG 1641

Db       1652 GAATGGGTGGCATG 1665
```

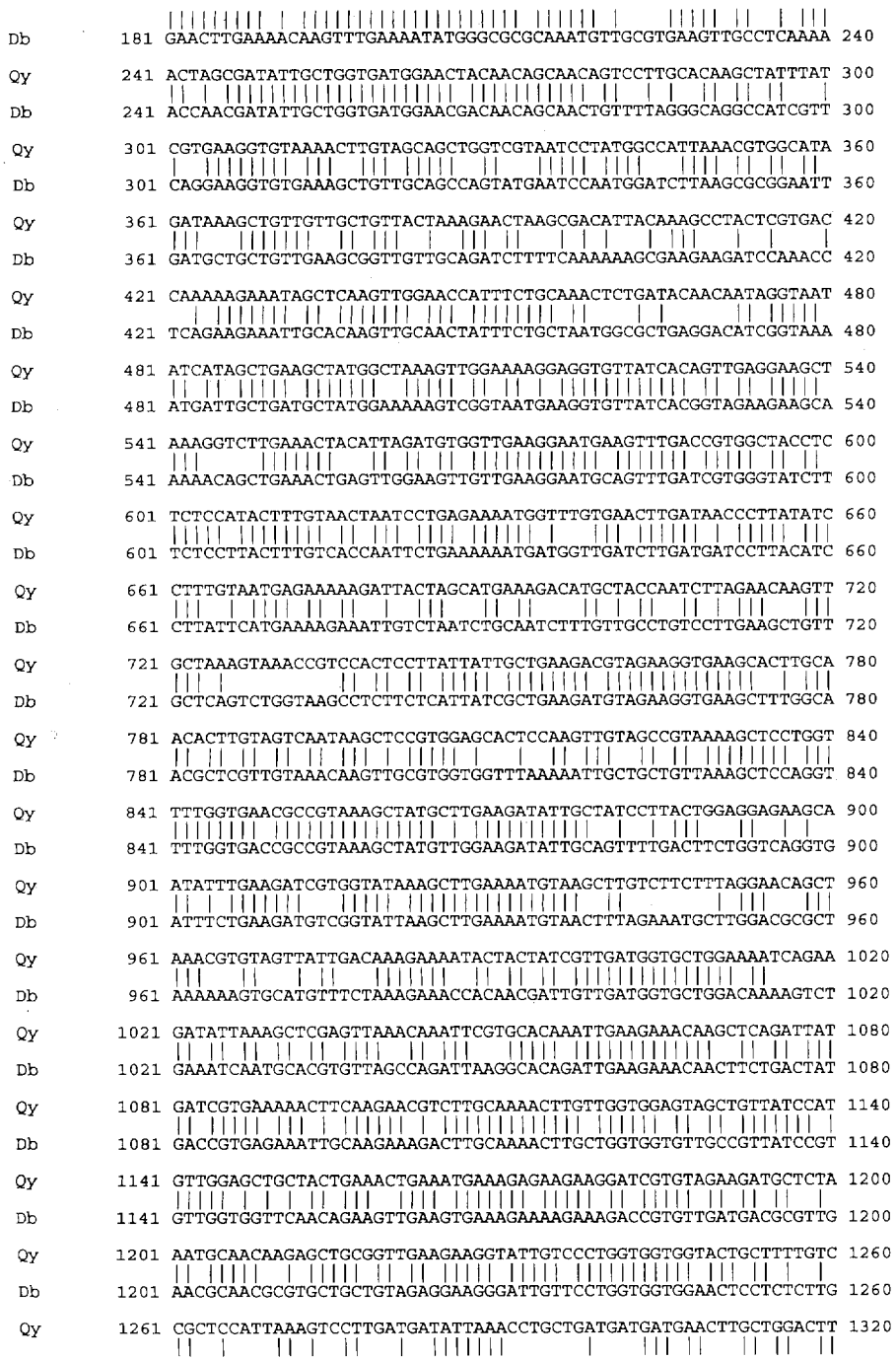
```
RESULT 5
AB057417
LOCUS       AB057417                2715 bp    DNA     linear   BCT 16-MAR-2001
DEFINITION  Pseudoalteromonas sp. PS1M3 gene for GroES, GroEL, complete cds.
ACCESSION   AB057417
VERSION     AB057417.1  GI:13366171
KEYWORDS    .
SOURCE      Pseudoalteromonas sp. PS1M3
  ORGANISM  Pseudoalteromonas sp. PS1M3
            Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
            Alteromonadaceae; Pseudoalteromonas.
REFERENCE   1  (sites)
  AUTHORS   Kurusu,Y. and Nakamura,T.
  TITLE     Moleculer chaperone of Psychrophile
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2715)
  AUTHORS   Kurusu,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2001) Yasurou Kurusu, Ibaraki University School
            of Agriculture, Laboratory of Molecular Microbiology; chuo 3-21-1,
            Inashiki Ami, Ibaraki 300-0393, Japan
            (E-mail:krsy@ipc.ibaraki.ac.jp, Tel:81-298-88-8646,
            Fax:81-298-88-8646)
FEATURES             Location/Qualifiers
     source          1. .2715
                     /organism="Pseudoalteromonas sp. PS1M3"
                     /mol_type="genomic DNA"
                     /strain="PS1M3"
                     /db_xref="taxon:87791"
                     /country="Japan"
                     /note="isolated from the Japan Trench"
     promoter        376. .388
     -35_signal      383. .388
     promoter        402. .410
     -10_signal      406. .411
     gene            458. .761
                     /gene="groES"
     RBS             458. .464
                     /gene="groES"
     CDS             474. .761
                     /gene="groES"
                     /codon_start=1
                     /transl_table=11
                     /product="GroES"
                     /protein_id="BAB39464.1"
                     /db_xref="GI:13366172"
                     /translation="MNIRPLQDRVIVKRLEEETKSAGGIVLTGSAAEKSTRGEVVAVG
                     NGRILDNGDVRALEVKAGDTVLFGSYVEKTEKIEGQEYLIMREDNILGIVG"
     gene            797. .2460
                     /gene="groEL"
     RBS             797. .803
                     /gene="groEL"
     CDS             811. .2460
                     /gene="groEL"
                     /function="molecular chaperone"
                     /codon_start=1
                     /transl_table=11
                     /product="GroEL"
                     /protein_id="BAB39465.1"
                     /db_xref="GI:13366173"
                     /translation="MAAKEVLFAGDARAKMLTGVNILANAVKVTLGPKGRNVVLDKSF
                     GSPVITKDGVSVAKEIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQSIVNE
                     GLKAVAAGMNPMDLKRGIDKAIIAAVAELKALSVPCADTKAIAQVGTISANSDKEIGD
                     IIAEAMEKVGRNTGVITVEEGQSLENELDVVEGMQFDRGYLSPYFINNAEKGAVELDN
                     PFILLVDKKVSNIRELLPTLEAVAKASKPLLIIAEDLEGEALATLVVNNMRGIVKVSA
                     VKAPGFGDRRKAMLQDIAVLTGGTVISEEIGLELEKATVEDLGTAKRVVITKDDTTII
                     DGAGEEEGITGRVSQIKAQIEEATSDYDKEKLQERMAKLAGGVAVIKVGAATEIEMKE
                     KKDRVEDALHATRAAVEEGVVPGGGVALVRAASKLVELLGDNEDQNHGVKVALRAMEA
                     PLRQIVTNAGDEASVVINAVKGGEGNYGYNAATGEYNDMIEMGILDPTKVTRSALQFA
                     GSIAGLMITTEAMVAEIPKEEAAGAPDMGGMGGMGGMGGMM"
     terminator      2519. .2567
BASE COUNT      860 a    513 c    613 g    729 t
ORIGIN
```

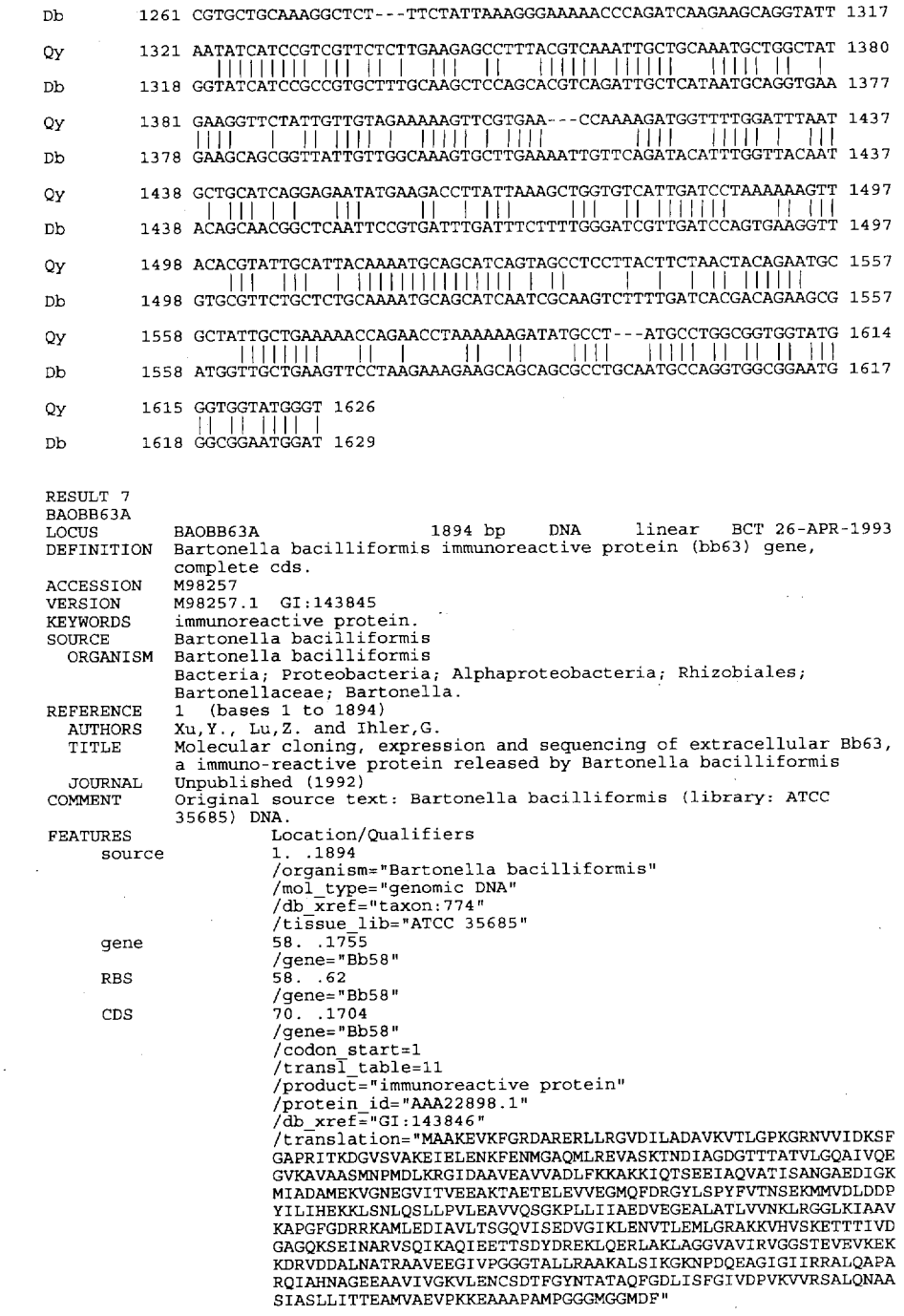
```
  Query Match             47.2%;  Score 777;  DB 1;  Length 2715;
  Best Local Similarity   67.4%;  Pred. No. 1.8e-128;
  Matches 1111;  Conservative    0;  Mismatches  535;  Indels    3;  Gaps    1;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db        811 ATGGCAGCTAAAGAAGTACTTTTTGCAGGTGACGCACGCGCTAAAATGCTAACTGGCGTA 870

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Db        871 AACATCCTAGCAAACGCAGTTAAAGTTACATTAGGTCCTAAAGGCCGTAACGTTGTACTA 930

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180

Db        931 GACAAATCGTTTGGCTCTCCAGTGATCACTAAAGATGGTGTATCTGTAGCAAAAGAAATC 990

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240

Db        991 GAACTTGAAGATAAGTTTGAAAACATGGGCGCGCAAATGGTTAAAGAAGTTGCATCTAAA 1050

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Db       1051 GCAAATGATGCAGCCGGCGATGGTACAACTACCGCTACAGTACTTGCACAGTCTATTGTT 1110

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Db       1111 AATGAAGGCCTTAAAGCGGTTGCTGCAGGCATGAACCCAATGGACCTTAAACGTGGTATT 1170

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db       1171 GATAAAGCAATCATTGCTGCTGTTGCAGAGCTTAAGGCGCTTTCTGTTCCTTGTGCTGAT 1230

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db       1231 ACTAAAGCTATTGCTCAAGTAGGTACTATTTCAGCTAACTCTGATAAAGAAATTGGCGAC 1290

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTG---GAAAAGGAGGTGTTATCACAGTTGAGGAA 537

Db       1291 ATCATTGCAGAAGCAATGGAAAAAGTAGGCCGTAATACAGGTGTTATTACTGTTGAAGAA 1350

Qy        538 GCTAAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTAC 597

Db       1351 GGTCAATCACTAGAGAACGAACTAGATGTTGTAGAAGGCATGCAGTTTGACCGTGGTTAC 1410
```

```
Qy        598 CTCTCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTAT 657
Db       1411 CTATCTCCTTACTTCATTAACAACGCTGAAAAAGGCGCAGTTGAATTAGATAACCCATTT 1470

Qy        658 ATCCTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAA 717
Db       1471 ATTCTACTTGTAGATAAAAAAGTATCTAACATTCGTGAGCTACTACCTACATTAGAAGCC 1530

Qy        718 GTTGCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTT 777
Db       1531 GTTGCTAAAGCAAGCAAGCCACTATTAATCATTGCTGAAGACCTTGAAGGCGAAGCACTT 1590

Qy        778 GCAACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCT 837
Db       1591 GCAACGCTTGTAGTTAACAACATGCGCGGCATTGTTAAGGTATCAGCGGTTAAAGCTCCT 1650

Qy        838 GGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAA 897
Db       1651 GGTTTTGGCGACCGTCGTAAAGCGATGTTACAAGATATCGCCGTATTAACTGGCGGTACT 1710

Qy        898 GCAATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACA 957
Db       1711 GTTATTTCTGAAGAAATTGGCCTTGAGCTTGAAAAAGCAACAGTTGAAGACCTAGGTACA 1770

Qy        958 GCTAAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCA 1017
Db       1771 GCTAAGCGCGTTGTTATCACTAAAGACGATACAACTATCATCGATGGTGCTGGCGAAGAA 1830

Qy       1018 GAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGAT 1077
Db       1831 GAAGGCATTACCGGTCGTGTTTCTCAAATTAAAGCACAAATTGAAGAAGCAACATCAGAC 1890

Qy       1078 TATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATC 1137
Db       1891 TACGATAAAGAAAAACTACAAGAGCGTATGGCTAAACTTGCCGGCGGTGTTGCAGTAATT 1950

Qy       1138 CATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCT 1197
Db       1951 AAAGTAGGCGCAGCTACTGAAATTGAAATGAAAGAGAAAAAAGACCGTGTTGAAGATGCA 2010

Qy       1198 CTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTT 1257
Db       2011 TTACATGCAACTCGTGCAGCGGTTGAAGAAGGTGTAGTACCGGGTGGCGGTGTTGCTCTA 2070

Qy       1258 GTCCGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db       2071 GTACGTGCAGCTAGCAAGCTAGTTGAACTACTTGGCGATAACGAAGATCAAAACCACGGT 2130

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db       2131 GTTAAAGTTGCCCTTCGTGCGATGGAAGCACCTCTTCGTCAAATCGTAACTAACGCGGGT 2190

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437
Db       2191 GACGAAGCATCTGTTGTTATTAACGCTGTTAAAGGTGGCGAAGGTAACTACGGTTACAAC 2250

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       2251 GCAGCAACTGGCGAATACAATGACATGATTGAAATGGGTATCCTAGATCCAACTAAAGTA 2310

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       2311 ACGCGCTCTGCACTACAATTTGCAGGGTCTATAGCGGGTCTAATGATCACAACTGAAGCA 2370

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db       2371 ATGGTTGCTGAAATCCCTAAAGAAGAAGCTGCTGGTGCCCCTGATATGGGTGGCATGGGC 2430

Qy       1618 GGTATGGGTGGTATGGACGGTATGTACTA 1646
Db       2431 GGAATGGGTGGCATGGGCGGCATGATGTA 2459
```

```
RESULT 6
BBGROELT
LOCUS       BBGROELT                1635 bp    DNA     linear   BCT 22-MAR-1995
DEFINITION  B.bacilliformis GroEL-type gene encoding heat shock protein 60.
ACCESSION   Z15160
VERSION     Z15160.1  GI:397595
KEYWORDS    groEL gene; heat shock protein; heat shock protein 60.
SOURCE      Bartonella bacilliformis
  ORGANISM  Bartonella bacilliformis
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bartonellaceae; Bartonella.
REFERENCE   1  (bases 1 to 1635)
  AUTHORS   Schreiber,M.G.
  TITLE     The 60 kD heat shock protein of Bartonella bacilliformis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1635)
  AUTHORS   Schreiber,M.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-1992) Michael G. Schreiber Dr., Virologie,
            Bernhard Nocht Institute for, Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg, Hamburg, D-2000 Hamburg 36,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1635
                     /organism="Bartonella bacilliformis"
                     /mol_type="genomic DNA"
                     /strain="Bartonella bacilliformis B13"
                     /db_xref="taxon:774"
                     /chromosome="5 kb EcoRI genomic DNA fragment"
                     /clone="7B2"
                     /clone_lib="Bar bac gt11"
     gene            1. .1635
                     /gene="GroEL-type"
     CDS             1. .1635
                     /gene="GroEL-type"
                     /standard_name="Bb HSP 60"
                     /note="heat shock protein 60 (HSP 60)"
                     /codon_start=1
                     /transl_table=11
                     /product="GroEL"
                     /protein_id="CAA78859.1"
                     /db_xref="GI:397596"
                     /db_xref="SWISS-PROT:P35635"
                     /translation="MAAKEVKFGRDARERLLRGVDILADAVKVTLGPKGRNVVIDKSF
                     GAPRITKDGVSVAKEIELENKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQE
                     GVKAVAASMNPMDLKRGIDAAVEAVVADLFKKAKKIQTSEEIAQVATISANGAEDIGK
                     MIADAMEKVGNEGVITVEEAKTAETELEVVEGMQFDRGYLSPYFVTNSEKMMVDLDDP
                     YILIHEKKLSNLQSLLPVLEAVAQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAV
                     KAPGFGDRRKAMLEDIAVLTSGQVISEDVGIKLENVTLEMLGRAKKVHVSKETTTIVD
                     GAGQKSEINARVSQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGSTEVEVKEK
                     KDRVDDALNATRAAVEEGIVPGGGTPLLRAAKALSIKGKNPDQEAGIGIIRRALQAPA
                     RQIAHNAGEEAAVIVGKVLENCSDTFGYNTATAQFRDLISFGIVDPVKVVRSALQNAA
                     SIASLLITTEAMVAEVPKKEAAAPAMPGGGMGGMDF"
BASE COUNT      471 a    283 c    425 g    456 t
ORIGIN
```

```
  Query Match           46.8%;  Score 771.2;  DB 1;  Length 1635;
  Best Local Similarity  68.3%;  Pred. No. 2.1e-127;
  Matches 1115;  Conservative    0;  Mismatches  508;  Indels    9;  Gaps    3;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 ATGGCTGCTAAAGAAGTAAAATTTGGCCGTGACGCACGTGAAAGGTTGTTACGTGGTGTG 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         61 GATATCCTTGCTGATGCTGTTAAGGTAACACTTGGTCCTAAAGGCCGCAATGTGGTGATT 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        121 GATAAATCATTTGGTGCTCCACGCATTACAAAAGATGGTGTGTCCGTTGCAAAAGAAATC 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
```





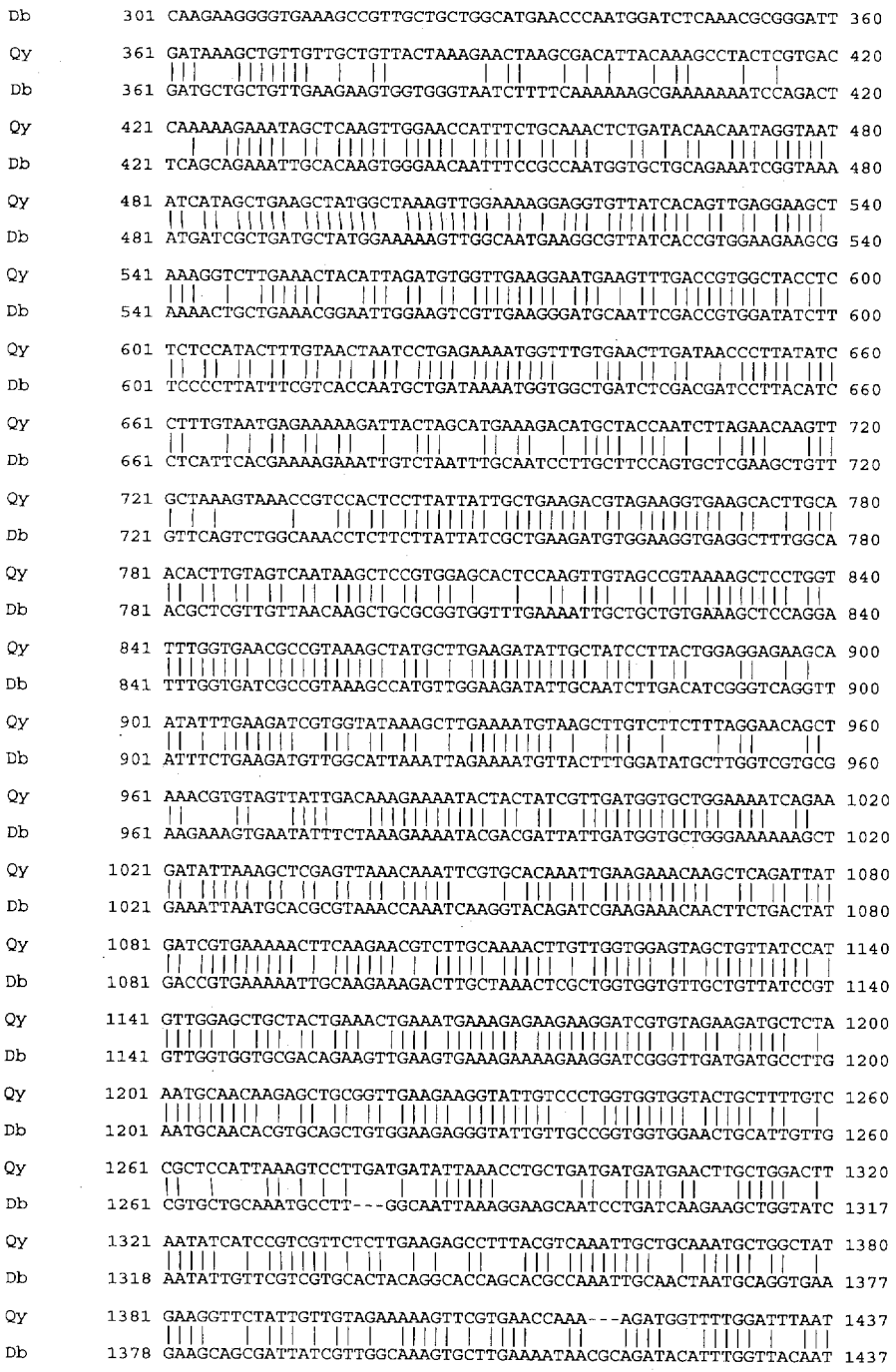
RESULT 7
BAOBB63A
LOCUS BAOBB63A 1894 bp DNA linear BCT 26-APR-1993
DEFINITION Bartonella bacilliformis immunoreactive protein (bb63) gene, complete cds.
ACCESSION M98257
VERSION M98257.1 GI:143845
KEYWORDS immunoreactive protein.
SOURCE Bartonella bacilliformis
ORGANISM Bartonella bacilliformis
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bartonellaceae; Bartonella.
REFERENCE 1 (bases 1 to 1894)
AUTHORS Xu,Y., Lu,Z. and Ihler,G.
TITLE Molecular cloning, expression and sequencing of extracellular Bb63, a immuno-reactive protein released by Bartonella bacilliformis
JOURNAL Unpublished (1992)
COMMENT Original source text: Bartonella bacilliformis (library: ATCC 35685) DNA.
FEATURES Location/Qualifiers
source 1. .1894
/organism="Bartonella bacilliformis"
/mol_type="genomic DNA"
/db_xref="taxon:774"
/tissue_lib="ATCC 35685"
gene 58. .1755
/gene="Bb58"
RBS 58. .62
/gene="Bb58"
CDS 70. .1704
/gene="Bb58"
/codon_start=1
/transl_table=11
/product="immunoreactive protein"
/protein_id="AAA22898.1"
/db_xref="GI:143846"
/translation="MAAKEVKFGRDARERLLRGVDILADAVKVTLGPKGRNVVIDKSF
GAPRITKDGVSVAKEIELENKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQE
GVKAVAASMNPMDLKRGIDAAVEAVVADLFKKAKKIQTSEEIAQVATISANGAEDIGK
MIADAMEKVGNEGVITVEEAKTAETELEVVEGMQFDRGYLSPYFVTNSEKMMVDLDDP
YILIHEKKLSNLQSLLPVLEAVVQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAV
KAPGFGDRRKAMLEDIAVLTSGQVISEDVGIKLENVTLEMLGRAKKVHVSKETTTIVD
GAGQKSEINARVSQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGSTEVEVKEK
KDRVDDALNATRAAVEEGIVPGGGTALLRAAKALSIKGKNPDQEAGIGIIRRALQAPA
RQIAHNAGEEAAVIVGKVLENCSDTFGYNTATAQFGDLISFGIVDPVKVVRSALQNAA
SIASLLITTEAMVAEVPKKEAAAPAMPGGGMGGMDF"

```
     terminator      1727. .1755
                     /gene="Bb58"
     stem_loop       1748. .1788
BASE COUNT      557 a    326 c    476 g    535 t
ORIGIN

  Query Match             46.8%;  Score 771.2;  DB 1;  Length 1894;
  Best Local Similarity   68.3%;  Pred. No. 2.1e-127;
  Matches 1115;  Conservative    0;  Mismatches  508;  Indels    9;  Gaps    3;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db         70 ATGGCTGCTAAAGAAGTAAAATTTGGCCGTGACGCACGTGAAAGGTTGTTACGTGGTGTG 129

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        130 GATATCCTTGCTGATGCTGTTAAGGTAACACTTGGTCCTAAAGGCCGCAATGTGGTGATT 189

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        190 GATAAATCATTTGGTGCTCCACGCATTACAAAAGATGGTGTGTCCGTTGCAAAAGAAATC 249

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        250 GAACTTGAAAACAAGTTTGAAAATATGGGCGCGCAAATGTTGCGTGAAGTTGCCTCAAAA 309

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        310 ACCAACGATATTGCTGGTGATGGAACGACAACAGCAACTGTTTTAGGGCAGGCCATCGTT 369

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        370 CAGGAAGGTGTGAAAGCTGTTGCAGCCAGTATGAATCCAATGGATCTTAAGCGCGGAATT 429

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        430 GATGCTGCTGTTGAAGCGGTTGTTGCAGATCTTTTCAAAAAAGCGAAGAAGATCCAAACC 489

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        490 TCAGAAGAAATTGCACAAGTTGCAACTATTTCTGCTAATGGCGCTGAGGACATCGGTAAA 549

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        550 ATGATTGCTGATGCTATGGAAAAAGTCGGTAATGAAGGTGTTATCACGGTAGAAGAAGCA 609

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        610 AAAACAGCTGAAACTGAGTTGGAAGTTGTTGAAGGAATGCAGTTTGATCGTGGGTATCTT 669

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        670 TCTCCTTACTTTGTCACCAATTCTGAAAAAATGATGGTTGATCTTGATGATCCTTACATC 729

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        730 CTTATTCATGAAAAGAAATTGTCTAATCTGCAATCTTTGTTGCCTGTGCTTGAAGCTGTT 789

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        790 GTTCAGTCTGGTAAGCCTCTTCTCATTATCGCTGAAGATGTAGAAGGTGAAGCTTTGGCA 849

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        850 ACGCTCGTTGTAAACAAGTTGCGTGGTGGTTTAAAAATTGCTGCTGTTAAAGCTCCAGGT 909

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        910 TTTGGTGACCGCCGTAAAGCTATGTTGGAAGATATTGCAGTTTTGACTTCTGGTCAGGTG 969

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        970 ATTTCTGAAGATGTCGGTATTAAGCTTGAAAATGTAACTTTAGAAATGCTTGGACGCGCT 1029

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       1030 AAAAAAGTGCATGTTTCTAAAGAAACCACAACCATTGTTGATGGTGCTGGACAAAAGTCT 1089

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1090 GAAATCAATGCACGTGTTAGCCAGATTAAGGCACAGATTGAAGAAACAACTTCTGACTAT 1149

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1150 GACCGTGAGAAATTGCAAGAAAGACTTGCAAAACTTGCTGGTGGTGTTGCCGTTATCCGT 1209

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1210 GTTGGTGGTTCAACAGAAGTTGAAGTGAAAGAAAAGAAAGACCGTGTTGATGACGCGTTG 1269

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1270 AACGCAACACGTGCTGCTGTAGAGGAAGGGATTGTTCCTGGTGGTGGAACTGCGCTCTTG 1329

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1330 CGTGCTGCAAAGGCTCT---TTCTATTAAAGGGAAAAACCCAGATCAAGAAGCAGGTATT 1386

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1387 GGTATCATCCGCCGTGCTTTGCAAGCTCCAGCACGTCAGATTGCTCATAATGCAGGTGAA 1446

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAA---CCAAAAGATGGTTTTGGATTTAAT 1437
Db       1447 GAAGCAGCGGTTATTGTTGGCAAAGTGCTTGAAAATTGTTCAGATACATTTGGTTACAAT 1506

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       1507 ACAGCAACGGCTCAATTCGGTGATTTGATTTCTTTTGGGATCGTTGATCCAGTGAAGGTT 1566

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       1567 GTGCGTTCTGCTCTGCAAAATGCAGCATCAATCGCAAGTCTTTTGATCACGACAGAAGCG 1626

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
Db       1627 ATGGTTGCTGAAGTTCCGAAGAAAGAAGCAGCAGCGCCTGCAATGCCAGGTGGCGGAATG 1686

Qy       1615 GGTGGTATGGGT 1626
Db       1687 GGCGGAATGGAT 1698

RESULT 8
BHU78514
LOCUS       BHU78514                1644 bp    DNA     linear   BCT 06-SEP-1997
DEFINITION  Bartonella henselae heat shock protein HSP60 (groEL) gene, complete
            cds.
ACCESSION   U78514
VERSION     U78514.1  GI:2358233
KEYWORDS    .
SOURCE      Bartonella henselae
  ORGANISM  Bartonella henselae
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bartonellaceae; Bartonella.
REFERENCE   1  (bases 1 to 1644)
  AUTHORS   Haake,D.A., Summers,T.A., McCoy,A.M. and Schwartzman,W.
  TITLE     Heat shock response and groEL sequence of Bartonella henselae and
            Bartonella quintana
  JOURNAL   Microbiology 143 (Pt 8), 2807-2815 (1997)
  MEDLINE   97419519
   PUBMED   9274034
REFERENCE   2  (bases 1 to 1644)
  AUTHORS   Haake,D.A., Summers,T.A. and Schwartzman,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Medicine, 111F, West Los Angeles VAMC,
```

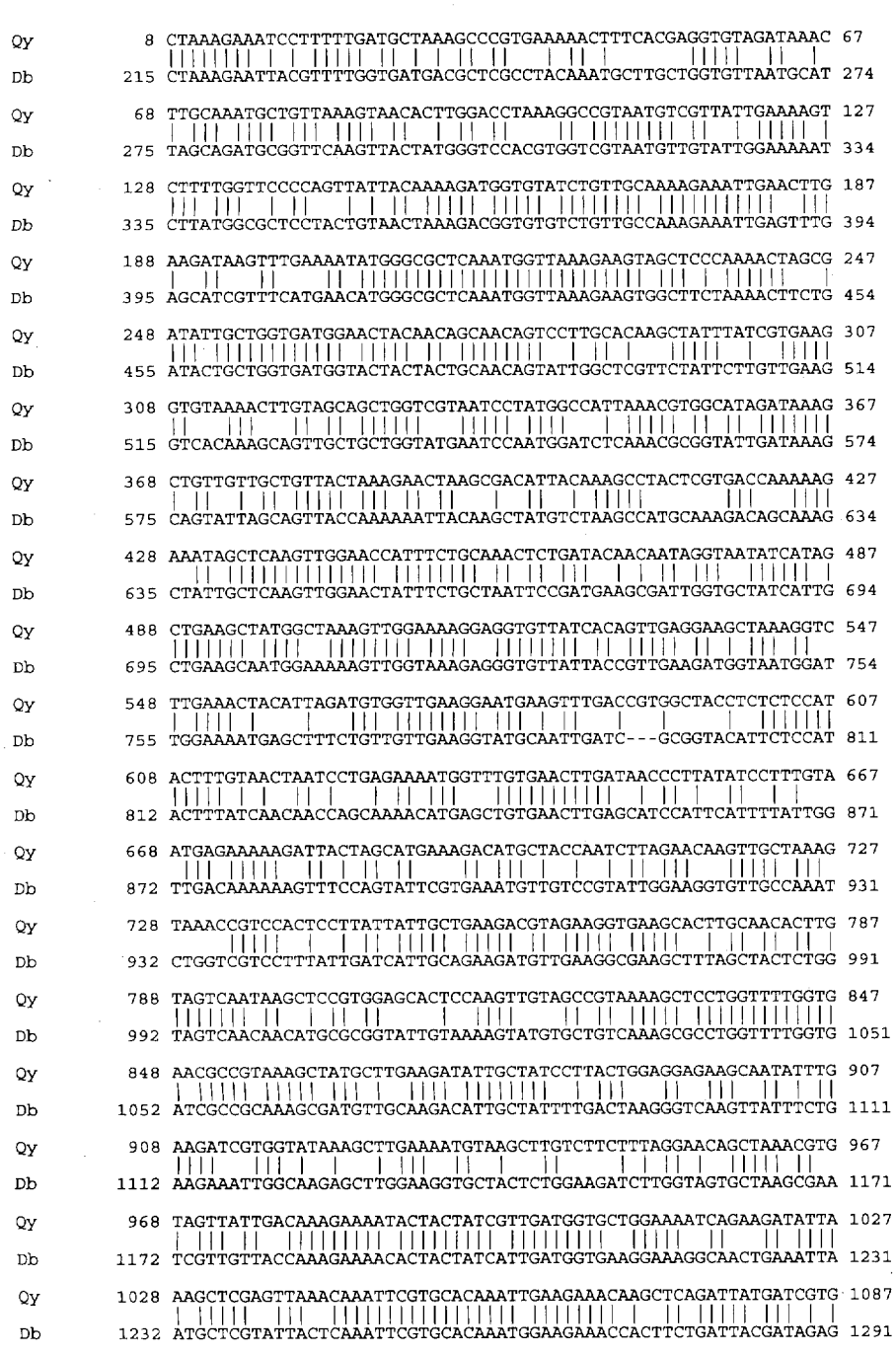
```
          11301 Wilshire Blvd., Los Angeles, CA 90073, USA
FEATURES             Location/Qualifiers
     source          1..1644
                     /organism="Bartonella henselae"
                     /mol_type="genomic DNA"
                     /strain="ATCC 49882"
                     /db_xref="taxon:38323"
     gene            1..1644
                     /gene="groEL"
     CDS             1..1644
                     /gene="groEL"
                     /note="GroEL"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein HSP60"
                     /protein_id="AAB69094.1"
                     /db_xref="GI:2358234"
                     /translation="MAAKEVKFGREARERLLRGVDILANAVKVTLGPKGRNVVIDKSF
                     GAPRITKDGVSVAKEIELEDKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQE
                     GVKAVAAGMNPMDLKRGIDAAVDEVVANLFKKAKKIQTSAEIAQVGTISANGAAEIGK
                     MIADAMEKVGNEGVITVEEAKTAETELEVVEGMQFDRGYLSPYFVTNAEKMVADLDDP
                     YILIHEKKLSNLQSLLPVLEAVVQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAV
                     KAPGFGDRRKAMLEDIAILTSGQVISEDVGIKLENVTLDMLGRAKKVNISKENTTIID
                     GAGQKSEINARVNQIKVQIEETTSDYDREKLQERLAKLAGGVAVIRVGGATEVEVKEK
                     KDRVDDALNATRAAVEEGIVAGGGTALLRAANALTVKGSNPDQEAGINIVRRALQAPA
                     RQIATNAGEEAAIIVGKVLENNADTFGYNTATGEFGDLIALGIVDPVKVVRSALQNAA
                     SIASLLITTEAMVAEVPKKDTPVPPMPGGGMGGMGGMDF"
BASE COUNT      486 a    309 c    419 g    430 t
ORIGIN

  Query Match             46.8%;  Score 770.6;  DB 1;  Length 1644;
  Best Local Similarity   68.1%;  Pred. No. 2.7e-127;
  Matches 1118;  Conservative    0;  Mismatches  514;  Indels    9;  Gaps    3;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 ATGGCTGCTAAAGAAGTGAAGTTTGGCCGTGAAGCGCGTGAGCGTTTGTTGCGCGGCGTT 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         61 GATATCCTTGCTAACGCTGTTAAAGTAACACTCGGCCCTAAAGGGCGCAATGTGGTGATC 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        121 GATAAATCATTTGGTGCGCCTCGCATTACAAAAGATGGTGTATCCGTTGCAAAAGAAATC 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        181 GAACTGGAAGATAAGTTCGAAAATATGGGTGCGCAAATGTTGCGTGAAGTTGCTTCTAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        241 ACCAATGACATTGCTGGTGATGGAACAACAACGGCAACTGTTTTAGGACAGGCTATTGTC 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        301 CAAGAAGGTGTAAAAGCCGTTGCTGCAGGCATGAACCCAATGGATCTCAAACGTGGTATC 360

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        361 GATGCTGCTGTTGATGAAGTGGTGGCAAATCTTTTCAAAAAAGCGAAAAAAATCCAAACT 420

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        421 TCAGCAGAAATTGCACAAGTGGGAACAATTTCCGCTAATGGTGCTGCTGAAATCGGCAAA 480

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        481 ATGATCGCTGATGCCATGGAAAAAGTTGGCAATGAAGGTGTCATTACCGTTGAAGAAGCA 540

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        541 AAAACTGCTGAAACGGAATTAGAAGTCGTTGAAGGAATGCAGTTTGATCGTGGATATCTT 600

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        601 TCCCCTTACTTTGTCACCAATGCTGAGAAAATGGTGGCTGATCTTGATGATCCTTACATT 660

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        661 CTTATTCACGAAAAGAAACTCTCTAATCTCCAATCCTTGCTTCCAGTACTTGAAGCTGTT 720

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        721 GTTCAGTCTGGTAAACCACTTCTCATTATCGCTGAAGATGTGGAAGGTGAAGCTTTGGCA 780

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        781 ACGCTCGTTGTCAACAAACTGCGTGGTGGTTTGAAAATTGCTGCTGTGAAAGCTCCAGGA 840

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        841 TTTGGTGATCGCCGTAAAGCCATGTTAGAAGATATTGCAATCTTAACATCCGGACAGGTT 900

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        901 ATTTCTGAAGATGTTGGCATTAAATTGGAAAATGTCACTTTGGATATGCTTGGTCGTGCA 960

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        961 AAGAAAGTCAATATTTCTAAAGAAAATACCACGATTATTGATGGTGCTGGACAAAAGAGC 1020

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1021 GAAATTAATGCGCGCGTCAACCAAATCAAGGTACAGATCGAAGAAACAACTTCTGACTAT 1080

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1081 GACCGTGAAAAATTGCAAGAAAGACTTGCTAAACTCGCTGGAGGTGTTGCTGTTATCCGT 1140

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1141 GTTGGTGGAGCAACAGAAGTTGAAGTGAAAGAAAAGAAAGATCGTGTTGATGATGCCTTG 1200

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1201 AACGCAACACGTGCAGCTGTGGAAGAAGGTATTGTTGCCGGTGGTGGAACCGCATTGTTG 1260

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1261 CGTGCAGCAAATGCCCTGACTG---TTAAAGGAAGCAATCCTGATCAGGAAGCTGGTATC 1317

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1318 AATATTGTTCGTCGTGCACTCCAAGCGCCAGCACGCCAAATTGCTACCAATGCCGGTGAA 1377

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAA---AGATGGTTTTGGATTTAAT 1437
Db       1378 GAAGCAGCGATTATCGTTGGCAAAGTGCTTGAAAACAATGCAGATACATTTGGTTACAAC 1437

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       1438 ACCGCAACCGGTGAATTTGGTGATTTGATTGCTTTGGGAATCGTTGATCCTGTGAAGGTT 1497

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       1498 GTGCGCTCTGCTCTTCAGAATGCTGCCTCAATTGCTAGCCTTCTCATTACAACAGAAGCA 1557

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
Db       1558 ATGGTTGCTGAAGTTCCAAAGAAAGACACACCAGTGCCTCCAATGCCTGGTGGCGGAATG 1617

Qy       1615 GGTGGTATGGGTGGTATGGAC 1635
Db       1618 GGTGGAATGGGCGGCATGGAC 1638
```

```
RESULT 9
PHA243594
LOCUS       PHA243594              2241 bp    DNA     linear   BCT 05-FEB-2001
DEFINITION  Pseudoalteromonas haloplanktis groEL, groES genes.
ACCESSION   AJ243594
VERSION     AJ243594.2  GI:12697197
KEYWORDS    groEL gene; GroEL protein; groES gene; GROES protein.
SOURCE      Pseudoalteromonas haloplanktis
  ORGANISM  Pseudoalteromonas haloplanktis
            Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
            Alteromonadaceae; Pseudoalteromonas.
REFERENCE   1
  AUTHORS   Tosco,A., Birolo,L., Scaloni,A., Sannia,G. and Marino,G.
  TITLE     A GroEL-like protein from the psychrotrophic bacterium
            Pseudoalteromonas haloplanktis TAC125
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Tosco,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Tosco A., Organic and Biological Chemistry,
            University of Naples, via Mezzocannone 16 (NA), I-80134, ITALY
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 2241)
  AUTHORS   Stefania,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Stefania M., Organic and Biological
            Chemistry, University of Naples, via Mezzocannone 16 (NA), I-80134,
            ITALY
COMMENT     On Feb 6, 2001 this sequence version replaced gi:5524757.
FEATURES             Location/Qualifiers
     source          1..2241
                     /organism="Pseudoalteromonas haloplanktis"
                     /mol_type="genomic DNA"
                     /strain="TAC125"
                     /db_xref="taxon:228"
     gene            71..133
                     /gene="groESL"
     -35_signal      71..83
                     /gene="groESL"
                     /evidence=experimental
     -10_signal      97..105
                     /gene="groESL"
                     /note="sigma 32-dependent promoter"
                     /evidence=experimental
     -35_signal      105..110
                     /gene="groESL"
                     /evidence=experimental
     -10_signal      128..133
                     /gene="groESL"
                     /note="sigma 70-dependent promoter"
                     /evidence=experimental
     gene            153..455
                     /gene="groES"
     RBS             153..156
                     /gene="groES"
     CDS             168..455
                     /gene="groES"
                     /function="molecular co-chaperone"
                     /codon_start=1
                     /transl_table=11
                     /product="groES protein"
                     /protein_id="CAC28359.1"
                     /db_xref="GI:12697198"
                     /translation="MNIRPLHDRVIVKRLEEETKSAGGIVLTGSAAEKSTRGEVVAVG
                     NGRILESGDVRALEVKAGDTVLFGSYVEKVEKIEGQEYLIMREDNILGIVG"
     gene            489..2147
                     /gene="groEL"
     RBS             489..493
                     /gene="groEL"
     CDS             504..2147
                     /gene="groEL"
                     /function="molecular chaperone"
                     /codon_start=1
                     /transl_table=11
                     /product="GroEL protein"
                     /protein_id="CAB50775.1"
                     /db_xref="GI:5524758"
                     /db_xref="SPTREMBL:Q9XAU7"
                     /translation="MAAKEVLFAGDARAKMLTGVNILANAVKVTLGPKGRNVVLDKSF
                     GSPVITKDGVSVAKEIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQSIVNE
                     GLKAVAAGMNPMDLKRGIDKAVIAAVAELKALSVPCSDTKAIAQVGTISANSDKEIGD
                     IIAQAMEKVGRNSGVITVEEGQSLENELDVVEGMQFDRGYLSPYFINSPEKGTVELDN
                     PFILLVDKKISNIRELLPTLEAVAKASKPLLIIAEDLEGEALATLVVNNMRGIVKVSA
                     VKAPGFGDRRKAMLQDIAVLTGGTVISEEIGLELEKATVEDLGTAERVIITKDDTTII
                     DGAGEEAGINGRVSQIKAQIEEATSDYDKEKLQERMAKLAGGVAVIKVGAATEMEMKE
                     KKDRVEDALNATRAAVEEGVVPGGGVALVRAASKLVDLVGDNEDQNHGIKVALRAMEA
                     PLRQIVTNAGDEASVVINAVKAGSGNFGYNAATGEYNDMIEMGILDPTKVTRSALQFA
                     GSIAGLMITTEAMVAEIPKDDSAPDMGGMGGMGGMGGMM"
     gene            2178..2202
                     /gene="groESL"
     terminator      2178..2202
                     /gene="groESL"
BASE COUNT      735 a    416 c    511 g    579 t
ORIGIN

  Query Match             46.4%;  Score 764.6;  DB 1;  Length 2241;
  Best Local Similarity   67.4%;  Pred. No. 3e-126;
  Matches 1111;  Conservative    0;  Mismatches  529;  Indels    9;  Gaps    2;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db        504 ATGGCAGCAAAAGAAGTACTTTTTGCAGGTGACGCACGCGCTAAAATGCTAACTGGCGTA 563

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        564 AACATCCTGGCAAACGCAGTAAAAGTTACATTAGGTCCTAAAGGCCGTAACGTTGTACTA 623

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        624 GATAAATCATTTGGCTCTCCAGTTATCACTAAAGATGGTGTATCTGTAGCAAAAGAAATC 683

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        684 GAACTTGAAGACAAGTTTGAGAACATGGGCGCACAAATGGTTAAAGAAGTTGCATCTAAA 743

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        744 GCTAACGATGCAGCCGGCGATGGTACAACTACCGCTACAGTACTTGCACAGTCTATTGTA 803

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        804 AACGAAGGCCTTAAAGCTGTTGCAGCAGGCATGAACCCAATGGACCTTAAGCGCGGCATC 863

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        864 GACAAAGCAGTAATTGCTGCTGTTGCAGAGCTTAAAGCGTTATCTGTTCCATGCTCTGAC 923

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        924 ACTAAAGCAATTGCACAAGTAGGTACTATTTCAGCTAACTCTGATAAAGAGATTGGCGAC 983

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTG---GAAAAGGAGGTGTTATCACAGTTGAGGAA 537
Db        984 ATCATTGCACAAGCAATGGAAAAAGTAGGCCGTAACAGTGGTGTTATTACTGTAGAAGAA 1043

Qy        538 GCTAAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTAC 597
Db       1044 GGTCAATCACTAGAAAATGAACTCGATGTTGTTGAAGGCATGCAGTTTGACCGTGGTTAC 1103

Qy        598 CTCTCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTAT 657
Db       1104 CTATCTCCTTACTTTATTAACAGCCCAGAAAAAGGTACTGTTGAACTAGATAACCCATTT 1163

Qy        658 ATCCTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAA 717
```

```
Db       1164 ATTCTACTTGTAGATAAAAAAATATCTAACATTCGCGAATTATTACCAACATTAGAAGCA 1223

Qy        718 GTTGCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTT 777

Db       1224 GTTGCTAAAGCAAGCAAGCCACTACTAATCATCGCAGAAGACCTTGAAGGCGAAGCATTA 1283

Qy        778 GCAACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCT 837

Db       1284 GCTACATTAGTAGTTAATAACATGCGTGGTATTGTTAAAGTATCAGCAGTGAAAGCTCCT 1343

Qy        838 GGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAA 897

Db       1344 GGTTTTGGCGACCGTCGTAAAGCAATGCTACAAGATATCGCTGTATTAACTGGCGGTACT 1403

Qy        898 GCAATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACA 957

Db       1404 GTGATTTCAGAAGAAATCGGCCTTGAGCTTGAAAAAGCAACAGTTGAAGACCTAGGTACA 1463

Qy        958 GCTAAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCA 1017

Db       1464 GCTGAGCGCGTAATTATCACTAAAGATGATACAACAATCATTGATGGTGCTGGCGAAGAA 1523

Qy       1018 GAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGAT 1077

Db       1524 GCAGGTATTAACGGCCGTGTTTCACAAATTAAAGCACAAATCGAAGAAGCAACATCAGAC 1583

Qy       1078 TATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATC 1137

Db       1584 TACGATAAAGAAAAACTACAAGAGCGCATGGCTAAACTAGCTGGCGGTGTTGCAGTAATC 1643

Qy       1138 CATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCT 1197

Db       1644 AAAGTTGGTGCTGCGACTGAAATGGAAATGAAAGAGAAAAAAGACCGCGTTGAAGATGCA 1703

Qy       1198 CTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTT 1257

Db       1704 TTAAATGCAACTCGTGCAGCGGTTGAAGAAGGTGTAGTACCTGGCGGCGGCGTTGCACTA 1763

Qy       1258 GTCCGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317

Db       1764 GTTCGTGCAGCAAGTAAGCTGGTTGATTTAGTTGGCGACAACGAAGATCAAAACCACGGT 1823

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377

Db       1824 ATTAAAGTTGCACTTCGTGCAATGGAAGCGCCACTTCGTCAAATCGTAACAAACGCAGGT 1883

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437

Db       1884 GACGAAGCTTCAGTTGTAATTAACGCAGTTAAAGCAGGTTCAGGTAACTTTGGTTACAAT 1943

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497

Db       1944 GCTGCTACTGGCGAATACAATGACATGATTGAAATGGGTATCCTAGATCCAACTAAAGTA 2003

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557

Db       2004 ACGCGCTCTGCACTACAATTTGCAGGGTCTATTGCGGGTCTAATGATCACTACCGAAGCT 2063

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617

Db       2064 ATGGTAGCTGAAA------TCCCTAAAGATGATAGCGCTCCTGATATGGGTGGCATGGGC 2117

Qy       1618 GGTATGGGTGGTATGGACGGTATGTACTA 1646

Db       2118 GGTATGGGTGGAATGGGCGGTATGATGTA 2146


RESULT 10
RQU78515
LOCUS       RQU78515                1644 bp    DNA     linear   BCT 06-SEP-1997
DEFINITION  Bartonella quintana heat shock protein HSP60 (groEL) gene, complete
            cds.
ACCESSION   U78515
VERSION     U78515.1  GI:2358235
KEYWORDS    .
SOURCE      Bartonella quintana
  ORGANISM  Bartonella quintana
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bartonellaceae; Bartonella.
REFERENCE   1  (bases 1 to 1644)
  AUTHORS   Haake,D.A., Summers,T.A., McCoy,A.M. and Schwartzman,W.
  TITLE     Heat shock response and groEL sequence of Bartonella henselae and
            Bartonella quintana
  JOURNAL   Microbiology 143 (Pt 8), 2807-2815 (1997)
  MEDLINE   97419519
   PUBMED   9274034
REFERENCE   2  (bases 1 to 1644)
  AUTHORS   Haake,D.A., Summers,T.A. and Schwartzman,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Medicine, 111F, West Los Angeles VAMC,
            11301 Wilshire Blvd., Los Angeles, CA 90073, USA
FEATURES             Location/Qualifiers
     source          1. .1644
                     /organism="Bartonella quintana"
                     /mol_type="genomic DNA"
                     /strain="ATCC 51694"
                     /db_xref="taxon:803"
     gene            1. .1644
                     /gene="groEL"
     CDS             1. .1644
                     /gene="groEL"
                     /note="GroEL"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein HSP60"
                     /protein_id="AAB69095.1"
                     /db_xref="GI:2358236"
                     /translation="MAAKEVKFGREARERLLRGVDILANAVKVTLGPKGRNVVIDKSF
                     GAPRITKDGVSVAKEIELEDKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQE
                     GVKAVAAGMNPMDLKRGIDAAVEEVVGNLFKKAKKIQTSAEIAQVGTISANGAAEIGK
                     MIADAMEKVGNEGVITVEEAKTAETELEVVEGMQFDRGYLSPYFVTNADKMVADLDDP
                     YILIHEKKLSNLQSLLPVLEAVVQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAV
                     KAPGFGDRRKAMLEDIAILTSGQVISEDVGIKLENVTLDMLGRAKKVNISKENTTIID
                     GAGKKAEINARVNQIKVQIEETTSDYDREKLQERLAKLAGGVAVIRVGGATEVEVKEK
                     KDRVDDALNATRAAVEEGIVAGGGTALLRAANALAIKGSNPDQEAGINIVRRALQAPA
                     RQIATNAGEEAAIIVGKVLENNADTFGYNTATGQFGDLIALGIVDPVKVVRSALQNAA
                     SIASLLITTEAMVAEVPKKDTPMPPMPGGGMGGMGGMDF"
BASE COUNT      478 a    298 c    433 g    435 t
ORIGIN

  Query Match           46.3%;  Score 762.6;  DB 1;  Length 1644;
  Best Local Similarity  67.8%;  Pred. No. 7.1e-126;
  Matches 1113;  Conservative    0;  Mismatches  519;  Indels    9;  Gaps    3;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db          1 ATGGCTGCTAAAGAAGTCAAGTTCGGCCGTGAAGCGCGTGAGCGTTTGTTGCGCGGCGTT 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Db         61 GATATCCTTGCTAACGCTGTTAAAGTAACGCTCGGCCCTAAAGGTCGCAATGTTGTAATT 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180

Db        121 GATAAATCGTTTGGTGCGCCTCGCATCACAAAAGATGGTGTATCTGTTGCAAAAGAAATC 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240

Db        181 GAACTTGAAGATAAGTTTGAAAATATGGGTGCGCAAATGTTGCGCGAAGTTGCTTCTAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Db        241 ACCAATGACATTGCTGGGGATGGAACAACAACTGCAACTGTTTTAGGACAGGCTATTGTA 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
```



```
RESULT 11
LPNHTPB
LOCUS       LPNHTPB                 2016 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Legionella pneumophila 58-kDa common antigen (htpB) gene, complete
            cds.
ACCESSION   M91673 M35149
VERSION     M91673.1  GI:149689
KEYWORDS    58-kDa common antigen; htpB gene.
SOURCE      Legionella pneumophila
  ORGANISM  Legionella pneumophila
            Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
            Legionellaceae; Legionella.
REFERENCE   1  (bases 1 to 2016)
  AUTHORS   Sampson,J.S., O'Connor,S.P., Holloway,B.P., Plikaytis,B.B.,
            Carlone,G.M. and Mayer,L.W.
  TITLE     Nucleotide sequence of htpB, the Legionella pneumophila gene
            encoding the 58-kilodalton (kDa) common antigen, formerly
            designated the 60-kDa common antigen
  JOURNAL   Infect. Immun. 58 (9), 3154-3157 (1990)
  MEDLINE   90354095
   PUBMED   2117582
COMMENT     Original source text: Legionella pneumophila (strain Philadelphia
            1, sub_species pneumophila) DNA.
FEATURES             Location/Qualifiers
     source          1. .2016
                     /organism="Legionella pneumophila"
                     /mol_type="genomic DNA"
                     /strain="Philadelphia 1"
                     /sub_species="pneumophila"
                     /db_xref="taxon:446"
     gene            211. .1854
                     /gene="htpB"
     CDS             211. .1854
                     /gene="htpB"
                     /codon_start=1
                     /transl_table=11
                     /product="58-kDa common antigen"
                     /protein_id="AAA25298.1"
                     /db_xref="GI:149690"
                     /translation="MAKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYG
                     APTVTKDGVSVAKEIEFEHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEG
                     HKAVAAGMNPMDLKRGIDKAVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAI
                     IAEAMEKVGKEGVITVEDGNGLENELSVVEGMQLIAVHSPYFINNQQNMSCELEHPFI
                     LLVDKKVSSIREMLSVLEGVAKSGRPLLIIAEDVEGEALATLVVNNMRGIVKVCAVKA
                     PGFGDRRKAMLQDIAILTKGQVISEEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGE
                     GKATEINARITQIRAQMEETTSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKA
                     RVEDALHATRAAVEEGIVAGGGVALIRAQKALDSLKGDNDDQNMGINILRRAIESPMR
                     QIVTNAGYEASVVVNKVAEHKDNYGFNAATGEYGDMVEMGILDPTKVTRMALQNAASV
                     ASLMLTTECMVADLPKKEEGVGAGDMGGMGGMGGMGGMM"
BASE COUNT      591 a    323 c    508 g    594 t
ORIGIN

  Query Match             45.9%;  Score 756;  DB 1;  Length 2016;
  Best Local Similarity   67.3%;  Pred. No. 1e-124;
  Matches 1100;  Conservative    0;  Mismatches  525;  Indels    9;  Gaps    2;
```



RESULT 12
AF014829
LOCUS AF014829 1627 bp DNA linear BCT 20-MAY-1999
DEFINITION Bartonella henselae 60 kDa heat shock protein (groEL) gene, partial cds.
ACCESSION AF014829
VERSION AF014829.1 GI:3603158
KEYWORDS .
SOURCE Bartonella henselae
ORGANISM Bartonella henselae
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bartonellaceae; Bartonella.
REFERENCE 1 (bases 1 to 1627)
AUTHORS Marston,E.L., Sumner,J.W. and Regnery,R.L.
TITLE Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (groEL) of Bartonella species
JOURNAL Int. J. Syst. Bacteriol. 49 Pt 3, 1015-1023 (1999)
MEDLINE 99354490
PUBMED 10425758
REFERENCE 2 (bases 1 to 1627)
AUTHORS Marston,E.L.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Viral and Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G13, Atlanta, GA 30333, USA
FEATURES Location/Qualifiers
source 1. .1627
/organism="Bartonella henselae"
/mol_type="genomic DNA"
/strain="Houston-1"
/db_xref="taxon:38323"
gene 111. .>1627

```
                /gene="groEL"
     CDS        111. .>1627
                /gene="groEL"
                /codon_start=1
                /transl_table=11
                /product="60 kDa heat shock protein"
                /protein_id="AAD04238.1"
                /db_xref="GI:3603159"
                /translation="MVIDKSFGAPRITKDGVSVAKEIELEDKFENMGAQMLREVASKT
                NDIAGDGTTTATVLGQAIVQEGVKAVAAGMNPMDLKRGIDAAVDEVVANLFKKAKKIQ
                TSAEIAQVGTISANGAAEIGKMIADAMEKVGNEGVITVEEAKTAETELEVVEGMQFDR
                GYLSPYFVTNAEKMVADLDDPYILIHEKKLSNLQSLLPVLEAVVQSGKPLLIIAEDVE
                GEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTSGQVISEDVGIKLENVT
                LDMLGRAKKVNISKENTTIIDGAGQKSEINARVNQIKVQIEETTSDYDREKLQERLAK
                LAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGTALLRAANALTVK
                GSNPDQEAGINIVRRALQAPARQIATNAGEEAAIIVGKVLENNADTFGYNTATGEFGD
                LIALGIVDPVKVVRSALQNAASIASLLITTEAMVAEVPKKDTPVPPMPGGGMGGMG"
BASE COUNT      482 a    306 c    413 g    426 t
ORIGIN

  Query Match            45.8%;  Score 755;  DB 1;  Length 1627;
  Best Local Similarity  68.1%;  Pred. No. 1.6e-124;
  Matches 1111;  Conservative    0;  Mismatches  510;  Indels   10;  Gaps    4;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 ATGGCTGCTAAAGAAGTCAAGTTTGGCCGTGAAGCGCGTGAGCGTTTGTTGCGCGGCGTT 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         61 GATATCCTTGCTAACGCTGTTAAAGTAACACTCGGCCCTAAAGG-CGCAATGTGGTGATC 119

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        120 GATAAATCATTTGGTGCGCCTCGCATTACAAAAGATGGTGTATCCGTTGCAAAAGAAATC 179

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        180 GAACTGGAAGATAAGTTCGAAAATATGGGTGCGCAAATGTTGCGTGAAGTTGCTTCTAAA 239

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        240 ACCAATGACATTGCTGGTGATGGAACAACAACGGCAACTGTTTTAGGACAGGCTATTGTC 299

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        300 CAAGAAGGTGTAAAAGCCGTTGCTGCAGGCATGAACCCAATGGATCTCAAACGTGGTATC 359

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        360 GATGCTGCTGTTGATGAAGTGGTGGCAAATCTTTTCAAAAAAGCGAAAAAAATCCAAACT 419

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        420 TCAGCAGAAATTGCACAAGTGGGAACAATTTCCGCTAATGGTGCTGCTGAAATCGGCAAA 479

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        480 ATGATCGCTGATGCCATGGAAAAAGTTGGCAATGAAGGTGTCATTACCGTTGAAGAAGCA 539

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        540 AAAACTGCTGAAACGGAATTAGAAGTCGTTGAAGGAATGCAGTTTGATCGTGGATATCTT 599

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        600 TCCCCTTACTTTGTCACCAATGCTGAGAAAATGGTGGCTGATCTTGATGATCCTTACATT 659

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        660 CTTATTCACGAAAAGAAACTCTCTAATCTCCAATCCTTGCTTCCAGTACTTGAAGCTGTT 719

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        720 GTTCAGTCTGGTAAACCACTTCTCATTATCGCTGAAGATGTGGAAGGTGAAGCTTTGGCA 779

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        780 ACGCTCGTTGTCAACAAACTGCGTGGTGGTTTGAAAATTGCTGCTGTGAAAGCTCCAGGA 839

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        840 TTTGGTGATCGCCGTAAAGCCATGTTAGAAGATATTGCAATCTTAACATCCGGACAGGTT 899

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        900 ATTTCTGAAGATGTTGGCATTAAATTGGAAAATGTCACTTTGGATATGCTTGGTCGTGCA 959

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        960 AAGAAAGTCAATATTTCTAAAGAAAATACCACGATTATTGATGGTGCTGGACAAAAGAGC 1019

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1020 GAAATTAATGCGCGCGTCAACCAAATCAAGGTACAGATCGAAGAAACAACTTCTGACTAT 1079

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1080 GACCGTGAAAAATTGCAAGAAAGACTTGCTAAACTCGCTGGAGGTGTTGCTGTTATCCGT 1139

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1140 GTTGGTGGAGCAACAGAAGTTGAAGTGAAAGAAAAGAAAGATCGTGTTGATGATGCCTTG 1199

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1200 AACGCAACACGTGCAGCTGTGGAAGAAGGTATTGTTGCCGGTGGTGGAACCGCATTGTTG 1259

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1260 CGTGCAGCAAATGCCCTGACTG---TTAAAGGAAGCAATCCTGATCAGGAAGCTGGTATC 1316

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1317 AATATTGTTCGTCGTGCACTCCAAGCGCCAGCACGCCAAATTGCTACCAATGCCGGTGAA 1376

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAA---AGATGGTTTTGGATTTAAT 1437
Db       1377 GAAGCAGCGATTATCGTTGGCAAAGTGCTTGAAAACAATGCAGATACATTTGGTTACAAC 1436

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       1437 ACCGCAACCGGTGAATTTGGTGATTTGATTGCTTTGGGAATCGTTGATCCTGTGAAGGTT 1496

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       1497 GTGCGCTCTGCTCTTCAGAATGCTGCCTCAATTGCTAGCCTTCTCATTACAACAGAAGCA 1556

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
Db       1557 ATGGTTGCTGAAGTTCCAAAGAAAGACACACCAGTGCCTCCAATGCCTGGTGGCGGAATG 1616

Qy       1615 GGTGGTATGGG 1625
Db       1617 GGTGGAATGGG 1627


RESULT 13
BAP439085
LOCUS       BAP439085             1653 bp    DNA     linear   BCT 12-JUN-2002
DEFINITION  Buchnera aphidicola (Thelaxes suberi) groEL gene.
ACCESSION   AJ439085
VERSION     AJ439085.1  GI:21425626
KEYWORDS    groEL gene; GroEL protein.
SOURCE      Buchnera aphidicola (Thelaxes suberi)
  ORGANISM  Buchnera aphidicola (Thelaxes suberi)
```

```
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Buchnera.
REFERENCE   1
  AUTHORS   Fares,M.A., Barrio,E., Sabater-Munoz,B. and Moya,A.
  TITLE     The evolution of the heat-shock protein GroEL from Buchnera, the
            primary endosymbiont of aphids, is governed by positive selection
  JOURNAL   Mol. Biol. Evol. (2002) In press
REFERENCE   2  (bases 1 to 1653)
  AUTHORS   Sabater-Munoz,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2002) Sabater-Munoz B., Insto. Cavanilles Biodiv
            & Biol Evol, Universidad de Valencia, P.O. Box 2085, Valencia
            E-46071, SPAIN
FEATURES             Location/Qualifiers
     source          1..1653
                     /organism="Buchnera aphidicola (Thelaxes suberi)"
                     /mol_type="genomic DNA"
                     /specific_host="Thelaxes suberi aphid"
                     /db_xref="taxon:98797"
                     /country="Spain:Valencia"
     gene            1..1653
                     /gene="groEL"
     CDS             1..1653
                     /gene="groEL"
                     /function="molecular chaperonin"
                     /codon_start=1
                     /transl_table=11
                     /product="GroEL protein"
                     /protein_id="CAD27797.1"
                     /db_xref="GI:21425627"
                     /translation="MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSF
                     GAPAITKDGVTVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQSIVTE
                     GLKAVAAGMNPMDLKRGVDKAVIAAVEELKCLSVPCSDFKAIAQVGTISANSDETVGK
                     LIAEAMEKVGKEGVITVEEGTGLQDELDVVEGMQFDRGYLSPYFINKPDTGSIELDNP
                     FILLADKKISNNREMLPPLESVAKAGNLLIIAEDVEGEALATLVVNTMRGIVKVAAVK
                     APGFGDRRKAMLQDIATLTGSTVISEEIGLELEKTTIEDLGHAKRVVINKDTTTIIDG
                     IGNESAIKGRVVQIRQQIEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKK
                     ARVEDALHATRAAVEEGVVAGGGVALIRVAGKISSLTGENEDQNVGIKVAIRAMESPL
                     RQIVINAGEEASVIANNVKAGKGNYGYNAYSEQYGDMIKMGILDPTKVTRSALQYAAS
                     VAGLMITTECMITDLPKSDSSDLSNAGNGGMGGGMGGMGGMM"
BASE COUNT      575 a    205 c    382 g    491 t
ORIGIN

  Query Match             45.7%;  Score 753;  DB 1;  Length 1653;
  Best Local Similarity   66.6%;  Pred. No. 3.7e-124;
  Matches 1093;  Conservative    0;  Mismatches  545;  Indels    3;  Gaps    1;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 ATGGCAGCTAAAGACGTAAAATTTGGAAATGATGCTCGCGTAAAAATGCTTCGTGGAGTA 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         61 AATGTACTAGCTGATGCAGTAAAAGTTACTTTAGGACCAAAAGGTCGTAATGTAGTACTT 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        121 GATAAATCTTTTGGTGCTCCTGCTATTACTAAAGATGGTGTAACTGTAGCTCGTGAAATC 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        181 GAATTAGAAGATAAATTTGAAAACATGGGAGCTCAGATGGTAAAGGAAGTAGCATCTAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        241 GCGAACGATGCAGCTGGTGATGGAACTACTACAGCTACTGTATTAGCACAATCTATTGTA 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        301 ACAGAGGGATTAAAGGCTGTTGCAGCTGGAATGAATCCAATGGATTTAAAAAGAGGGGTA 360

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        361 GATAAAGCTGTTATTGCAGCAGTAGAAGAACTTAAATGTCTTTCTGTTCCTTGTTCTGAT 420

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        421 TTTAAAGCTATTGCTCAAGTTGGTACTATATCTGCAAATTCTGATGAAACTGTAGGAAAA 480

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        481 TTAATTGCAGAAGCAATGGAAAAGGTTGGAAAAGAAGGTGTAATCACTGTTGAAGAAGGA 540

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        541 ACTGGATTACAGGATGAATTAGATGTAGTAGAAGGAATGCAATTTGATCGTGGTTACTTA 600

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        601 TCTCCTTATTTCATTAATAAACCAGATACTGGATCTATAGAATTAGATAATCCATTTATT 660

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        661 CTTTTAGCAGATAAAAAGATTTCTAATAATCGAGAAATGCTTCCTCCTTTAGAATCTGTT 720

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        721 GCTAAAGCA---GGAAACCTATTAATTATTGCTGAAGATGTTGAAGGTGAAGCATTAGCA 777

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        778 ACATTAGTAGTAAATACTATGCGTGGTATAGTAAAAGTAGCTGCGGTAAAAGCACCTGGT 837

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        838 TTTGGTGATAGACGTAAAGCTATGCTACAAGATATTGCAACTTTAACAGGTAGCACAGTA 897

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        898 ATTTCTGAAGAAATTGGATTAGAATTAGAAAAAACTACAATAGAAGATTTAGGTCACGCT 957

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        958 AAAAGAGTAGTGATTAATAAAGATACTACTACTATTATTGATGGTATTGGTAATGAATCT 1017

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1018 GCTATTAAAGGTCGTGTTGTGCAAATACGTCAACAAATAGAAGAAGCAACATCTGATTAT 1077

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1078 GATCGTGAAAAATTACAAGAACGTGTAGCTAAACTTGCTGGTGGTGTAGCTGTTATTAAA 1137

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1138 GTTGGTGCTGCAACTGAAGTAGAAATGAAAGAGAAAAAAGCTCGTGTTGAAGATGCTTTA 1197

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1198 CATGCAACTCGTGCAGCTGTAGAAGAAGGTGTAGTTGCTGGTGGTGGTGTAGCATTAATT 1257

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1258 CGTGTTGCTGGTAAAATTAGTAGTTTAACAGGTGAAAATGAAGATCAAAATGTTGGAATT 1317

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1318 AAAGTAGCTATACGTGCAATGGAATCTCCACTTCGTCAAATAGTAATAAATGCAGGTGAA 1377

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       1378 GAAGCATCTGTTATAGCAAATAATGTAAAAGCTGGAAAAGGAAATTACGGTTATAATGCA 1437

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       1438 TATTCTGAACAATATGGTGATATGATTAAAATGGGTATTTTAGACCCAACTAAAGTTACT 1497
```

```
Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       1498 AGATCTGCTTTACAATATGCTGCATCTGTAGCTGGTTTAATGATTACTACTGAGTGTATG 1557

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db       1558 ATTACTGATCTACCTAAAAGCGATTCATCAGATTTAAGCAATGCTGGAAATGGTGGTATG 1617

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
Db       1618 GGTGGTGGTATGGGTGGCATG 1638


RESULT 14
LMGROELSR
LOCUS       LMGROELSR               2770 bp    DNA     linear   BCT 18-FEB-1992
DEFINITION  L.micdadei DNA for gro ELS operon.
ACCESSION   X57520
VERSION     X57520.1  GI:44099
KEYWORDS    groELS operon.
SOURCE      Tatlockia micdadei
  ORGANISM  Tatlockia micdadei
            Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
            Legionellaceae; Tatlockia.
REFERENCE   1  (bases 1 to 2770)
  AUTHORS   Hindersson,P., Hoiby,N. and Bangsborg,J.
  TITLE     Sequence analysis of the Legionella micdadei groELS operon
  JOURNAL   FEMS Microbiol. Lett. 77, 31-38 (1991)
FEATURES             Location/Qualifiers
     source          1. .2770
                     /organism="Tatlockia micdadei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:451"
BASE COUNT      846 a    520 c    675 g    729 t
ORIGIN

  Query Match             45.6%;  Score 751.2;  DB 1;  Length 2770;
  Best Local Similarity   67.6%;  Pred. No. 7e-124;
  Matches 1104;  Conservative    0;  Mismatches  518;  Indels   12;  Gaps    3;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db        472 CTAAAGAATTACGTTTTGGCGACGATGCTCGTCAACAAATGCTTGCTGGTGTAAACGCAT 531

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db        532 TGGCTGATGCAGTTAAAGCAACAATGGGTCCAAGCGGACGTAACGTTGTTTTAGAAAGAT 591

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        592 CATTTGGTGCACCTACTGTAACTAAAGACGGTGTTTCTGTTGCAAAAGAAATTGAATTCG 651

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        652 AAAACCGTTTCAAAAATATGGGCGCACAAATGGTTAAAGAAGTTGCTGCTAAAACTTCTG 711

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        712 ATACCGCAGGTGATGGTACTACTACTGCTACGGTATTGGCACGTGCCATTGTCGTTGAAG 771

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        772 GCCACAAAGCAGTTGCAGCAGGCATGAACCCAATGGATTTGAAACGAGGTATTGATAAAG 831

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        832 CAGTTGCCGCAGTTACCAAAAAATTACAAGAAATGTCTAAGCCTTGCAAAGACGGTAAAG 891

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        892 CAATTGCACAAGTAGGTACTATTTCTGCGAACTCTGATCAAGCAATCGGTTCTATCATTG 951

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        952 CTGAAGCAATGGAAAAAGTAGGTAAAGAAGGTGTTATTACTGTAGAAGATGGTAACAGCC 1011

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db       1012 TTGAAAACGAATTAGCTGTTGTTGAAGGTATGCAGTTTGATCGTGGTTACATTTCTCCAT 1071

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db       1072 ACTTCATCAACAATCAGCAAAACATGAGTGCTGAGCTTGAGCATCCTTTCATCCTCTTGG 1131

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db       1132 TTGACAAGAAAATTTCAACAATTCGTGACATGTTGTCTGTATTGGAAGCAGTTGCCAAAT 1191

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db       1192 CAGGACCGTCTTTATTGATCATTGCTGAAGATGTTGAAGGTGAAGCATTAGCAACTCTGG 1251

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db       1252 TTGTTAACAACATGCGCGGTATTGTTAAAGTGTGCGCGGTTAAAGCACCTGGTTTCGGTG 1311

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db       1312 ATCGTCGTAAAGCCATGTTACAAGACATTGCTATCCTTACTGCGGGCGAAGTCATTTCAG 1371

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db       1372 AAGAAGTGGGTACTAGTTTAGAAAGTGCTACTTTAGACAGCCTAGGCACTGCAAAACGCG 1431

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db       1432 TTGTCGTTACCAAAGAAAACACCACAATCATTGACGGTGAAGGTAAAGCGGCTGACATCA 1491

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1492 ATGCGCGTATTACTCAAATTCGTGCTCAAATGGAAGAAACCACTTCTGATTACGATCGCG 1551

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1552 AAAAATTACAAGAGCGCGTAGCTAAATTAGCTGGCGGTGTTGCCGTCAT-TAAGTTGGTG 1610

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1611 CTGCTACCGAATC-----GAATGAAGAGAAAAGCTCGTGTAGAAGACGCTCTTCATGCAA 1665

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1666 CTCGTGCTGCAGTAGAAGAAGGCATTGTTGCTGGTGGTGGTGTTGCCCTGATTCGTGCGC 1725

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1726 AAAAAGCATTAGATGGCTTGAAAGGCGAAAATGCTGACCAAGACATGGGTATTAACATCC 1785

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1786 TACGTCGCGCCATCGAATCTCCTCTCCGCCAAATCGTAGCCAATGCAGGTTATGAGCCTT 1845

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1846 CTGTTATCGTTAACAAAGTAGCTGAAAGCAAAGATAACTTTGGTTTCAACGCAGCTACTG 1905

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1906 GCGAATACGGTGACATGGTTGAGATGGGTATTCTTGATCCAACTAAAGTGACTCGTACTG 1965

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       1966 CATTACAAAACGCAGCTTCTGTTGCAAGCTTAATGCTGACCACCGAGTGCATGGTTGCTG 2025

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627
```

```
Db       2026 A------TCTACCTAAGAAAGAAGAACCTATGGGTGCCGGCGAAATGGGCGGCATGGGTG 2079

Qy       1628 GTATGGACGGTATG 1641
Db       2080 GAATGGGCGGCATG 2093


RESULT 15
AX607355
LOCUS       AX607355                1623 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 5284 from Patent WO02092818.
ACCESSION   AX607355
VERSION     AX607355.1  GI:28402886
KEYWORDS    .
SOURCE      Streptococcus agalactiae
  ORGANISM  Streptococcus agalactiae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1
  AUTHORS   Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
            Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
            Kunst,F.
  TITLE     Streptococcus agalactiae genome sequence, use for developing
            vaccines, diagnostic tools, and for identifying therapeutic targets
  JOURNAL   Patent: WO 02092818-A 5284 21-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..1623
                     /organism="Streptococcus agalactiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1311"
BASE COUNT      499 a    271 c    391 g    462 t
ORIGIN

  Query Match             45.3%;  Score 746.6;  DB 6;  Length 1623;
  Best Local Similarity   68.0%;  Pred. No. 5e-123;
  Matches 1072;  Conservative    0;  Mismatches  499;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db          5 CAAAAGATATTAAATTTTCAGCAGATGCCCGCTCAGCAATGGTGCGTGGTGTTGATATTT 64

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db         65 TAGCTGATACAGTCAAAGTAACATTAGGTCCTAAAGGCCGTAATGTTGTTCTTGAAAAAG 124

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        125 CATTTGGTTCTCCTTTAATTACAAATGATGGTGTGACAATTGCTAAAGAAATTGAGCTAG 184

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        185 AAGATCACTTTGAAAATATGGGAGCTAAACTTGTGTCAGAAGTGGCTTCAAAAACTAATG 244

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        245 ATATTGCAGGGGATGGCACTACAACTGCTACTGTTTTGACCCAAGCTATTGTACGGGAAG 304

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        305 GTCTTAAAAAATGTAACTGCAGGGGCAAATCCGATTGGCATTCGTCGTGGTATTGAAACAG 364

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        365 CTGTTTCAGCAGCAGTTGAAGAGCTAAAAGAGATTGCACAACCAGTTTCAGGCAAAGAAG 424

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        425 CTATTGCTCAAGTTGCGGCTGTGTCTTCACGTTCTGA---AAAAGTTGGGGAATACATTT 481

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        482 CTGAAGCTATGGGGCGCGTGGGTAATGATGGTGTTATCACTATTGAAGAATCGCGAGGTA 541

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db        542 TGGAAACAGAGCTTGAAGTTGTGGAAGGAATGCAGTTTGACCGTGGGTACTTGTCACAGT 601

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db        602 ATATGGTAACTGATAACGAGAAAATGGTCTCTGAACTTGAGAATCCGTATATCCTTATTA 661

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db        662 CAGATAAGAAAATTTCAAATATCCAAGAAATTTTACCATTATTAGAAGAGGTTCTTAAAA 721

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db        722 CAAATCGTCCGTTGCTAATCATCGCTGATGATGTTGATGGAGAAGCTCTCCCAACGCTTG 781

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db        782 TTCTTAACAAAATTCGTGGAACTTTCAATGTCGTAGCTGTTAAAGCGCCTGGATTTGGTG 841

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db        842 ATCGTCGTAAAGCCATGCTGGAAGATATTGCTATCCTAACAGGAGGAACTGTCGTTACTG 901

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db        902 AAGACCTTGGTTTAGACTTAAAAGATGCTACTATGCAAGTTTTAGGACAGTCTGCTAAAG 961

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db        962 TAACAGTAGATAAAGATTCTACTGTTATTGTCGAAGGTGCCGGTGACTCATCAGCAATTG 1021

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1022 CTAATCGCGTAGCTATCATTAAGTCACAGATGGAGGCTACAACTTCTGATTTTGATCGTG 1081

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1082 AAAAATTACAAGAACGACTTGCTAAGTTAGCCGGTGGTGTAGCAGTAATTAAAGTTGGTG 1141

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1142 CAGCGACTGAAACAGAATTAAAAGAGATGAAACTTCGCATCGAAGATGCGTTAAATGCAA 1201

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1202 CGCGTGCTGCAGTTGAAGAAGGTATTGTTTCAGGTGGAGGTACGGCTCTTGTGAACGTTA 1261

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1262 TTGAAAAAGTAGCGGCACTGAAACTTAATGGTGATGAGGA---GACTGGACGTAATATTG 1318

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1319 TTCTTCGTGCTCTCGAAGAGCCTGTTCGTCAAATTGCTTACAATGCTGGATATGAAGGTT 1378

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1379 CAGTTATTATTGAACGTTTAAAACAGTCTGAAATTGGTACAGGATTTAATGCGGCCAATG 1438

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1439 GAGAATGGGTAGATATGGTTACCACAGGTATCATTGACCCTGTCAAAGTAACACGTTCTG 1498

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       1499 CACTTCAAAATGCGGCATCTGTAGCAAGTCTTATCTTGACTACAGAAGCAGTAGTAGCAA 1558

Qy       1568 AAAAACCAGAACCTAAA 1584
```

Db 1559 ATAAACCTGAACCAGAA 1575

Search completed: January 29, 2004, 14:44:06
Job time : 4112 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 09:58:37 ; Search time 352 Seconds
(without alignments)
12630.609 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 1647
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1647	100.0	1647	18	AAT69201	Lawsonia intracell
2	1647	100.0	4972	18	AAT69203	Lawsonia intracell
3	746.6	45.3	1620	24	ABN69159	Streptococcus poly
c 4	746.6	45.3	2155561	24	ABN71527	Streptococcus poly
5	741.8	45.0	96109	22	AAF28548	Genomic fragment #
6	732.2	44.5	1901	24	ABQ70705	Listeria monocytog
7	729	44.3	213251	24	ABQ67193	Listeria innocua c
c 8	729	44.3	3011208	24	ABQ69245	Listeria innocua D
c 9	722.6	43.9	2944528	24	ABA03041	Listeria monocytog
10	721	43.8	1830121	17	AAT42063	Haemophilus influe
11	713.6	43.3	2465	12	AAQ13136	Hyp operon. Chlam
12	710.4	43.1	2223	12	AAQ13137	Hyp operon. Chlam
13	710	43.1	640681	24	ABA92787	Buchnera sp. genom
14	708.4	43.0	2365589	24	ABA90521	Genomic sequence o
15	704.2	42.8	1661	20	AAX86155	DNA encoding a Str
16	701	42.6	1635	24	ABN69160	Streptococcus poly
17	698.4	42.4	3625	20	AAX12979	Enterococcus faeca
18	698.4	42.4	3625	24	ABS98774	Enterococcus faeca
c 19	690.2	41.9	1230025	20	AAX91990	Nucleotide sequenc
20	689.8	41.9	1654	20	AAX86153	DNA encoding a Str
21	689.8	41.9	1926	22	AAF25036	Nucleotide sequenc
22	689.6	41.9	1635	24	ABL92607	Chlamydia pneumoni
23	689.2	41.8	2107	22	AAH56860	Antibiotic resista
24	687.6	41.7	1620	25	ABX07694	S. pneumoniae type
25	687.6	41.7	1647	22	AAH90906	2CFE 104 coding se
26	687.6	41.7	5365	19	AAV52210	Streptococcus pneu
c 27	687.6	41.7	2162598	25	ABS56454	Streptococcus pneu
28	686	41.7	1623	22	AAH90800	CFE 104 coding seq
29	686	41.7	1038602	20	AAZ01425	Complete genome se
30	684.4	41.6	1647	22	AAC90805	Escherichia coli G
31	684.4	41.6	2155	24	AAD33423	E. coli groESL gen
32	682.8	41.5	1647	21	AAA48498	E. coli groEL codi
33	682.8	41.5	1647	24	ABL96324	Nucleotide sequenc
34	682.8	41.5	1741	22	AAH56858	E. coli K12 strain
35	682.8	41.5	2032	22	AAH56856	Multidrug resistan
36	682.8	41.5	4524	21	AAA48500	Escherichia coli g
37	679.6	41.3	2401	22	AAH56859	S. pneumoniae groE
38	675.2	41.0	1570	24	ABQ69121	Listeria monocytog
39	673.2	40.9	1647	24	ABQ81662	Detrimental organi
40	671.6	40.8	1647	24	ABQ81665	Detrimental organi
41	670	40.7	1647	24	ABQ81654	Detrimental organi
42	670	40.7	1647	24	ABQ81663	Detrimental organi
43	670	40.7	1647	24	ABQ81664	Detrimental organi
44	667	40.5	2006	22	AAH56857	Uropathogen E. col
45	656	39.8	1638	24	ABQ81658	Detrimental organi

ALIGNMENTS

RESULT 1
AAT69201
ID AAT69201 standard; DNA; 1647 BP.
XX
AC AAT69201;
XX
DT 20-AUG-1997 (first entry)
XX
DE Lawsonia intracellularis GroEL DNA.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine;
KW GroEL; heat shock protein; ss.
XX
OS Lawsonia intracellularis.
XX
PN WO9720050-A1.
XX
PD 05-JUN-1997.
XX
PF 29-NOV-1996; 96WO-AU00767.
XX
PR 30-NOV-1995; 95AU-0006911.
PR 30-NOV-1995; 95AU-0006910.
XX
PA (DARA-) DARATECH PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Hasse D, Panaccio M;
XX
DR WPI; 1997-310605/28.
DR P-PSDB; AAW16678.

XX
PT Vaccine for treating or preventing Lawsonia intracellularis
PT infection - especially in pigs, containing non-pathogenic form of
PT bacterium or its components
XX
PS Claim 12; Page 38-42; 94pp; English.
XX
CC A DNA molecule (AAT69201) codes for the GroEL heat shock protein
CC (AAW16678) of Lawsonia intracellularis, the causative agent of porcine
CC proliferative enteropathy (PPE). A genomic library was prepd. from
CC L. intracellularis PPE lesion isolates and screened with rabbit
CC anti-L. intracellularis antiserum. Phagemid DNA was isolated from
CC individual clones and sequenced. GroEL and GroES (see also AAT69202)
CC sequences were identified and another 13 clones were subsequently
CC (AAT69203-15) isolated. These nucleic acids can be used to produce
CC recombinant polypeptides useful in vaccines against intestinal
CC diseases such as PPE, including recombinant vaccines utilising
CC bacterial, fungal or viral vectors. They can also be used as
CC genetic vaccines and in diagnostic assays.
XX
SQ Sequence 1647 BP; 536 A; 277 C; 368 G; 466 T; 0 other;

Query Match 100.0%; Score 1647; DB 18; Length 1647;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620

Qy       1621 ATGGGTGGTATGGACGGTATGTACTAG 1647
              ||||||||||||||||||||||||||||
Db       1621 ATGGGTGGTATGGACGGTATGTACTAG 1647
```

RESULT 2
AAT69203
ID AAT69203 standard; DNA; 4972 BP.
XX
AC AAT69203;
XX
DT 20-AUG-1997 (first entry)

XX
DE Lawsonia intracellularis vaccine candidate DNA.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine; ss.
XX
OS Lawsonia intracellularis.
XX
PN WO9720050-A1.
XX
PD 05-JUN-1997.
XX
PF 29-NOV-1996; 96WO-AU00767.
XX
PR 30-NOV-1995; 95AU-0006911.
PR 30-NOV-1995; 95AU-0006910.
XX
PA (DARA-) DARATECH PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Hasse D, Panaccio M;
XX
DR WPI; 1997-310605/28.
XX
PT Vaccine for treating or preventing Lawsonia intracellularis
PT infection - especially in pigs, containing non-pathogenic form of
PT bacterium or its components
XX
PS Claim 14; Page 47-50; 94pp; English.
XX
CC DNA molecules (AAT69203-15) were isolated from an immunoscreening of
CC a Lawsonia intracellularis library using experimental sera from
CC vaccinated pigs. These nucleic acids, as well as isolated GroEL
CC and GroES (AAT69201-02) sequences, encode putative vaccine candidates
CC (see also AAW16678-85) useful for protection of animals and birds
CC against intestinal diseases, esp. protection of pigs against
CC porcine proliferative enteropathy (PPE). They can also be used as
CC genetic vaccines.
XX
SQ Sequence 4972 BP; 1717 A; 871 C; 831 G; 1532 T; 21 other;

Query Match 100.0%; Score 1647; DB 18; Length 4972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 2107

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 2167

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 2227

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 2287

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 2347

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2348 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 2407

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2408 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 2467

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2468 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 2527

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2528 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 2587

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2588 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 2647

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2648 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 2707

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2708 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 2767

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2768 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 2827

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2828 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 2887

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2888 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 2947

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2948 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 3007

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3008 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 3067

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3068 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 3127

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3128 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 3187

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3188 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 3247

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3248 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 3307

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3308 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 3367

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3368 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 3427

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3428 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 3487

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3488 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 3547

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3548 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 3607

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3608 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 3667

Qy       1621 ATGGGTGGTATGGACGGTATGTACTAG 1647
             ||||||||||||||||||||||||||||
Db       3668 ATGGGTGGTATGGACGGTATGTACTAG 3694


RESULT 3
ABN69159
ID   ABN69159 standard; DNA; 1620 BP.
XX
AC   ABN69159;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 6231.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP28528.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3785; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 1620 BP; 498 A; 271 C; 390 G; 461 T; 0 other;

  Query Match             45.3%;  Score 746.6;  DB 24;  Length 1620;
  Best Local Similarity   68.0%;  Pred. No. 2.8e-162;
  Matches 1072;  Conservative    0;  Mismatches  499;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
              | ||||| ||   |||   ||  | ||||  |  | |    || |||| ||||
Db          5 CAAAAGATATTAAATTTTCAGCAGATGCCCGCTCAGCAATGGTGCGTGGTGTTGATATTT 64

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
              | ||  || | ||  |||||||| | || |||||||||||||||| ||| |||||||
Db         65 TAGCTGATACAGTCAAAGTAACATTAGGTCCTAAAGGCCGTAATGTTGTTCTTGAAAAAG 124

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
              | |||||||| ||  | |||||||| |||||||  |  |||| |||||||||||  || |
Db        125 CATTTGGTTCTCCTTTAATTACAAATGATGGTGTGACAATTGCTAAAGAAATTGAGCTAG 184

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
              ||||| | |||||||||||||| ||| || | ||  ||||| ||| | |||||||  |
Db        185 AAGATCACTTTGAAAATATGGGAGCTAAACTTGTGTCAGAAGTGGCTTCAAAAACTAATG 244

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
              ||||||| || ||||| |||||||| || || ||  |  | |||||||||   || ||||
Db        245 ATATTGCAGGGGATGGCACTACAACTGCTACTGTTTTGACCCAAGCTATTGTACGGGAAG 304

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
              || | |||  |||| | || ||   ||||| || | ||||  ||||| || || | ||
Db        305 GTCTTAAAAATGTAACTGCAGGGGCAAATCCGATTGGCATTCGTCGTGGTATTGAAACAG 364

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
              |||||  || |  | |||| ||||  || ||| || || ||  |  | | || |||||
Db        365 CTGTTTCAGCAGCAGTTGAAGAGCTAAAAGAGATTGCACAACCAGTTTCAGGCAAAGAAG 424

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
              || |||||||||  | | ||| ||  |||||  || | | || |  |||
Db        425 CTATTGCTCAAGTTGCGGCTGTGTCTTCACGTTCTGA---AAAAGTTGGGGAATACATTT 481

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
              ||||||||||||     || || || |  |||||||||||  |||| ||| |  ||||
Db        482 CTGAAGCTATGGAGCGCGTGGGTAATGATGGTGTTATCACTATTGAAGAATCGCGAGGTA 541

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
              | |||||    | || || |||||||||| |||||||||||||| ||| | || | |
Db        542 TGGAAACAGAGCTTGAAGTTGTGGAAGGAATGCAGTTTGACCGTGGGTACTTGTCACAGT 601

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
              | | |||||| ||   |||||||||| | ||||||||| || || |||||||||  ||
Db        602 ATATGGTAACTGATAACGAGAAAATGGTCTCTGAACTTGAGAATCCGTATATCCTTATTA 661

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
              || || || ||| | |  || |||| ||  ||||| | |||||| | |||  ||||
Db        662 CAGATAAGAAAATTTCAAATATCCAAGAAATTTTACCATTATTAGAAGAGGTTCTTAAAA 721

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
              ||| ||||| | || || || ||||| || || || || ||||| || |||| ||||
Db        722 CAAATCGTCCGTTGCTAATCATCGCTGATGATGTTGATGGAGAAGCTCTCCCAACGCTTG 781

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
              | | || ||  | |||||| |  || | || ||||| || ||||| ||||| |||||||
Db        782 TTCTTAACAAAATTCGTGGAACTTTCAATGTCGTAGCTGTTAAAGCGCCTGGATTTGGTG 841

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
              | || |||||||| |||||  ||||||||||||||||| || ||||||   |  |  ||
Db        842 ATCGTCGTAAAGCCATGCTGGAAGATATTGCTATCCTAACAGGAGGAACTGTCGTTACTG 901

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
              |||| | |||| || |  || ||| |  ||  |||||||    ||      |
Db        902 AAGACCTTGGTTTAGACTTAAAAGATGCTACTATGCAAGTTTTAGGACAGTCTGCTAAAG 961

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
              ||   | || |||||  |||| ||| || || ||||| ||  | |||  ||  |||
Db        962 TAACAGTAGATAAAGATTCTACTGTTATTGTCGAAGGTGCCGGTGACTCATCAGCAATTG 1021

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
              ||| ||       |||   |||| || || |  ||||  || |||| ||||||||
```

```
Db       1022 CTAATCGCGTAGCTATCATTAAGTCACAGATGGAGGCTACAACTTCTGATTTTGATCGTG 1081

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147

Db       1082 AAAAATTACAAGAACGACTTGCTAAGTTAGCCGGTGGTGTAGCAGTAATTAAAGTTGGTG 1141

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207

Db       1142 CAGCGACTGAAACAGAATTAAAAGAGATGAAACTTCGCATCGAAGATGCGTTAAATGCAA 1201

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267

Db       1202 CGCGTGCTGCAGTTGAAGAAGGTATTGTTTCAGGTGGAGGTACGGCTCTTGTGAACGTTA 1261

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327

Db       1262 TTGAAAAAGTAGCGGCACTGAAACTTAATGGTGATGAGGA---GACTGGACGTAATATTG 1318

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387

Db       1319 TTCTTCGTGCTCTCGAAGAGCCTGTTCGTCAAATTGCTTACAATGCTGGATATGAAGGTT 1378

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447

Db       1379 CAGTTATTATTGAACGTTTAAAACAGTCTGAAATTGGTACAGGATTTAATGCGGCCAATG 1438

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507

Db       1439 GAGAATGGGTAGATATGGTTACCACAGGTATCATTGACCCTGTCAAAGTAACACGTTCTG 1498

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567

Db       1499 CACTTCAAAATGCGGCATCTGTAGCAAGTCTTATCTTGACTACAGAAGCAGTAGTAGCAA 1558

Qy       1568 AAAAACCAGAACCTAAA 1584

Db       1559 ATAAACCTGAACCAGAA 1575
```

RESULT 4
ABN71527/c
ID ABN71527 standard; DNA; 2155561 BP.
XX
AC ABN71527;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 10967.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus sp.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 8; Page 4196-4488; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Query Match 45.3%; Score 746.6; DB 24; Length 2155561;
Best Local Similarity 68.0%; Pred. No. 2.6e-161;
Matches 1072; Conservative 0; Mismatches 499; Indels 6; Gaps 2;

```
Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67

Db    2024265 CAAAAGATATTAAATTTTCAGCAGATGCCCGCTCAGCAATGGTGCGTGGTGTTGATATTT 2024206

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127

Db    2024205 TAGCTGATACAGTCAAAGTAACATTAGGTCCTAAAGGCCGTAATGTTGTTCTTGAAAAAG 2024146

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187

Db    2024145 CATTTGGTTCTCCTTTAATTACAAATGATGGTGTGACAATTGCTAAAGAAATTGAGCTAG 2024086

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247

Db    2024085 AAGATCACTTTGAAAATATGGGAGCTAAACTTGTGTCAGAAGTGGCTTCAAAAACTAATG 2024026

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307

Db    2024025 ATATTGCAGGGGATGGCACTACAACTGCTACTGTTTTGACCCAAGCTATTGTACGGGAAG 2023966

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367

Db    2023965 GTCTTAAAAATGTAACTGCAGGGGCAAATCCGATTGGCATTCGTCGTGGTATTGAAACAG 2023906

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427

Db    2023905 CTGTTTCAGCAGCAGTTGAAGAGCTAAAAGAGATTGCACAACCAGTTTCAGGCAAAGAAG 2023846

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487

Db    2023845 CTATTGCTCAAGTTGCGGCTGTGTCTTCACGTTCTGA---AAAAGTTGGGGAATACATTT 2023789

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547

Db    2023788 CTGAAGCTATGGAGCGCGTGGGTAATGATGGTGTTATCACTATTGAAGAATCGCGAGGTA 2023729

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607

Db    2023728 TGGAAACAGAGCTTGAAGTTGTGGAAGGAATGCAGTTTGACCGTGGGTACTTGTCACAGT 2023669

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667

Db    2023668 ATATGGTAACTGATAACGAGAAAATGGTCTCTGAACTTGAGAATCCGTATATCCTTATTA 2023609
```

```
Qy       668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db   2023608 CAGATAAGAAAATTTCAAATATCCAAGAAATTTTACCATTATTAGAAGAGGTTCTTAAAA 2023549

Qy       728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db   2023548 CAAATCGTCCGTTGCTAATCATCGCTGATGATGTTGATGGAGAAGCTCTCCCAACGCTTG 2023489

Qy       788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db   2023488 TTCTTAACAAAATTCGTGGAACTTTCAATGTCGTAGCTGTTAAAGCGCCTGGATTTGGTG 2023429

Qy       848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db   2023428 ATCGTCGTAAAGCCATGCTGGAAGATATTGCTATCCTAACAGGAGGAACTGTCGTTACTG 2023369

Qy       908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db   2023368 AAGACCTTGGTTTAGACTTAAAAGATGCTACTATGCAAGTTTTAGGACAGTCTGCTAAAG 2023309

Qy       968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db   2023308 TAACAGTAGATAAAGATTCTACTGTTATTGTCGAAGGTGCCGGTGACTCATCAGCAATTG 2023249

Qy      1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db   2023248 CTAATCGCGTAGCTATCATTAAGTCACAGATGGAGGCTACAACTTCTGATTTTGATCGTG 2023189

Qy      1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db   2023188 AAAAATTACAAGAACGACTTGCTAAGTTAGCCGGTGGTGTAGCAGTAATTAAAGTTGGTG 2023129

Qy      1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db   2023128 CAGCGACTGAAACAGAATTAAAAGAGATGAAACTTCGCATCGAAGATGCGTTAAATGCAA 2023069

Qy      1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db   2023068 CGCGTGCTGCAGTTGAAGAAGGTATTGTTTCAGGTGGAGGTACGGCTCTTGTGAACGTTA 2023009

Qy      1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db   2023008 TTGAAAAAGTAGCGGCACTGAAACTTAATGGTGATGAGGA---GACTGGACGTAATATTG 2022952

Qy      1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db   2022951 TTCTTCGTGCTCTCGAAGAGCCTGTTCGTCAAATTGCTTACAATGCTGGATATGAAGGTT 2022892

Qy      1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db   2022891 CAGTTATTATTGAACGTTTAAAACAGTCTGAAATTGGTACAGGATTTAATGCGGCCAATG 2022832

Qy      1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db   2022831 GAGAATGGGTAGATATGGTTACCACAGGTATCATTGACCCTGTCAAAGTAACACGTTCTG 2022772

Qy      1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db   2022771 CACTTCAAAATGCGGCATCTGTAGCAAGTCTTATCTTGACTACAGAAGCAGTAGTAGCAA 2022712

Qy      1568 AAAAACCAGAACCTAAA 1584
Db   2022711 ATAAACCTGAACCAGAA 2022695
```

```
RESULT 5
AAF28548
ID   AAF28548 standard; DNA; 96109 BP.
XX
AC   AAF28548;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #35.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 345-368; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;
```

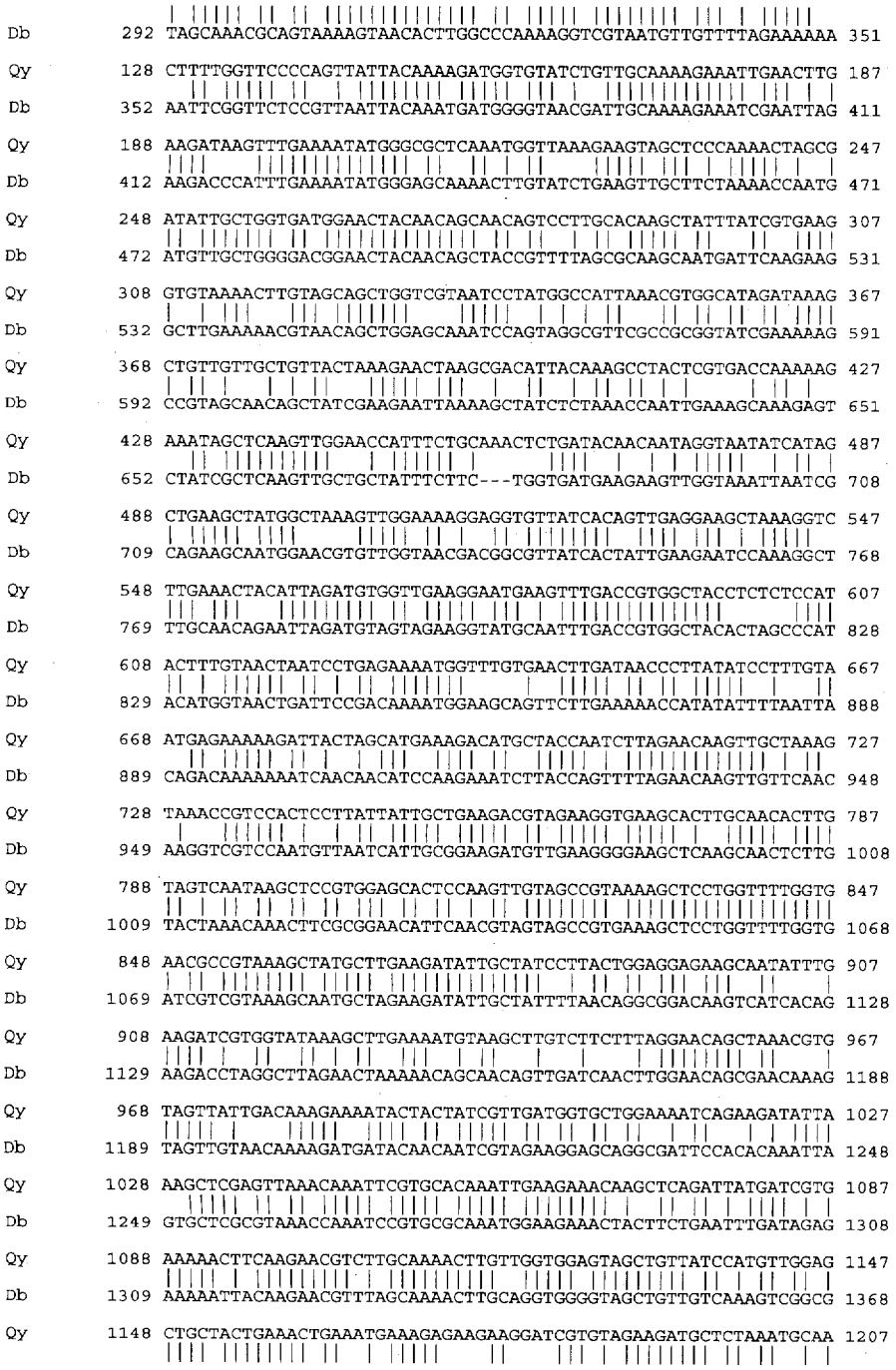
```
  Query Match           45.0%;  Score 741.8;  DB 22;  Length 96109;
  Best Local Similarity   66.3%;  Pred. No. 1.3e-160;
  Matches 1086;  Conservative    0;  Mismatches  547;  Indels    6;  Gaps    1;

Qy         8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db     49280 CAAAAGATGTAAGTTTTGGCTCAAATGCCCGCGAAAAAATGATTGATGGTGTTAACATTT 49339

Qy        68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db     49340 TGGCGGATGCTGTCAAGGTTACCCTAGGTCCAAAAGGCCGCAATGTGGTGATTGATAAAT 49399

Qy       128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db     49400 CATTTGGTGCGCCGACCATCACCAAAGATGGCGTGTCAGTTGCCAAAGAGATTGAACTTG 49459

Qy       188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db     49460 AAGATAAATTTGAAAACATGGGTGCACAATTGGTGCGTGAAGTTGCTTCTAAGACCAATG 49519

Qy       248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db     49520 ATGTGGCAGGCGATGGAACAACAACCGCAACAGTACTTGCTCAAGCCATCTTGGTTGAGG 49579

Qy       308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db     49580 GCATGAAAACGGTTGCTGCAGGCATGAATCCAATGGATCTAAAGCGTGGTATTGATAAAG 49639

Qy       368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db     49640 CCGTACGCGCCGCTGTTGAAGAAATCCGTGCTATCTCGACACCTGCTAATGACCATAAAG 49699
```

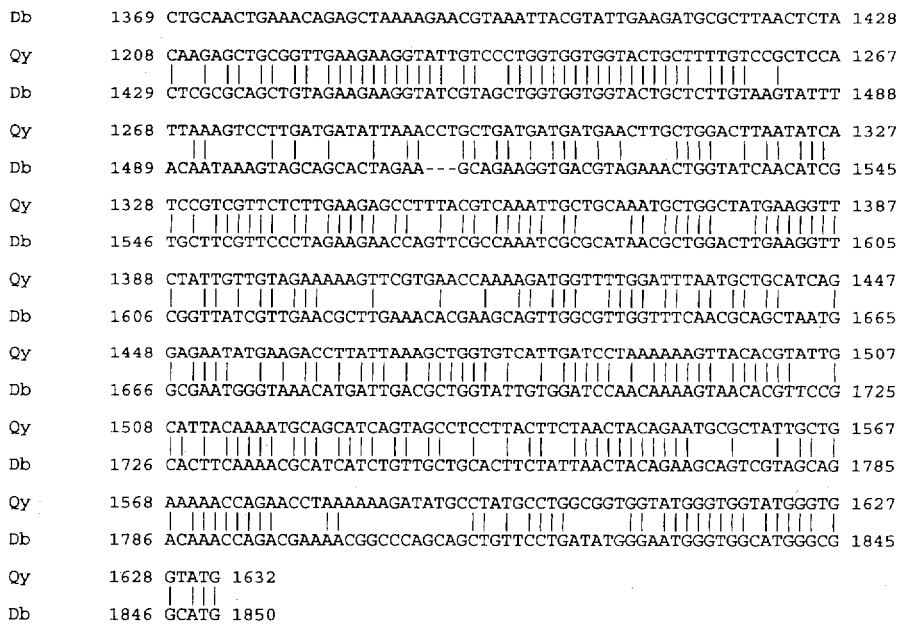
```
Qy       428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db     49700 CGATTGCTCAAGTTGGCTCAATTTCAGCAAACTCAGATGCTACCATTGGTGAGCTTATCT 49759

Qy       488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db     49760 CCAAGGCAATGGAAACAGTTGGCAAACAAGGCGTTATCACAGTTGAAGAGGGTTCAGGTT 49819

Qy       548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db     49820 TTGAAGATGCGCTTGAAGTAGTTGAGGGCATGCAGTTTGACCGTGGCTATATCAGCCCTT 49879

Qy       608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db     49880 ACTTTGCCAATAAACAAGACAGCTTGACTTGTGAATTTGACAATCCATTTATTCTTTTGG 49939

Qy       668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db     49940 TCGATAAAAAAATCTCAAATATTCGTGAGATTGTGCCACTACTTGAAAAAGTGATGCAAA 49999

Qy       728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db     50000 CCAGCCGTCCGCTATTAATCATTGCTGAAGATGTTGAAAATGAAGCATTGGCAACATTGG 50059

Qy       788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db     50060 TTGTCAATACTTTGCGTGGCGGATTAAAAACTTGTGCGGTTAAAGCGCCAGGTTTTGGTG 50119

Qy       848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db     50120 ATCGCCGTAAAGCCATGCTTCAAGACATTGCAATTTTAACAGGTGGCGTGGTTATCTCAG 50179

Qy       908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db     50180 AAGAAGTGGGCTTAAGCCTTGAGACTGCTGAGATTGAGCATTTGGGTACTGCCAAAAAAG 50239

Qy       968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db     50240 TAACCATTGGTAAAGAAAACACCGTCATTGTTGATGGTGCAGGTGATAAAGCCAGCATTG 50299

Qy      1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db     50300 AGGCGCGTGTAGAGTCAATTCGTCGTCAGGTTGAAGAATCAACTTCTGACTACGACAAAG 50359

Qy      1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db     50360 AAAAACTTCAAGAGCGTGTGGCAAAACTATCAGGCGGTGTTGCAGTCATCAAGGTGGGTG 50419

Qy      1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db     50420 CGGCAACTGAAACTGAGATGAAAGAGAAAAAAGACCGTGTAGACGATGCACTTCACGCTA 50479

Qy      1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db     50480 CCCGTGCTGCTGTTGAAGAGGGTGTTGTGCCTGGTGGTGGTGTTGCACTTGTGCGTGCTT 50539

Qy      1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db     50540 TATCTGCATTAAGCGACCTAAAAGGTGATAATGAAGACCAAAATGCAGGCATCAATATCC 50599

Qy      1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db     50600 TACGCCGTGCGATGGAAGCACCACTTCGCCAAATCGTCAGTAACGCAGGTGATGAGGCTT 50659

Qy      1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db     50660 CTGTAATTGTTAATGAAGTTAAAAATGGCTCTGGTAACTACGGTTATAATGCTGCATCAG 50719

Qy      1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db     50720 GTGAGTATGGCGATATGCTTGAAATGGGTATTTTAGATCCTGCCAAAGTAACCCGTTCAG 50779

Qy      1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db     50780 CATTAGAACATGCTGCTTCTGTCGCAGGTTTGATGTTGACCACTGAGGTGATGATTACCG 50839

Qy      1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627
Db     50840 ATAAACCAGCACC------AGAAGCACCAATGCCAGCAGGTGGTATGGGTGGTATGGGTG 50893

Qy      1628 GTATGGACGGTATGTACTA 1646
Db     50894 GTATGGGCGGTATGATGTA 50912
```

RESULT 6
ABQ70705
ID ABQ70705 standard; DNA; 1901 BP.
XX
AC ABQ70705;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #647.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 3518; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1901 BP; 660 A; 345 C; 435 G; 461 T; 0 other;

Query Match 44.5%; Score 732.2; DB 24; Length 1901;
Best Local Similarity 66.5%; Pred. No. 6.1e-159;
Matches 1081; Conservative 0; Mismatches 538; Indels 6; Gaps 2;

```
Qy         8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db       232 CAAAAGATATTAAATTTAGTGAAGATGCTCGTCGTGCCATGTTACGTGGTGTCGACCAAT 291

Qy        68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
```





RESULT 7
ABQ67193
ID ABQ67193 standard; DNA; 213251 BP.
XX
AC ABQ67193;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #6.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 5; SEQ ID 6; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes

```
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 213251 BP; 70939 A; 36026 C; 43257 G; 63029 T; 0 other;

  Query Match             44.3%;  Score 729;  DB 24;  Length 213251;
  Best Local Similarity   66.4%;  Pred. No. 1.5e-157;
  Matches 1079;  Conservative    0;  Mismatches  540;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
              | ||||| ||     ||| ||   | || |||        | | ||| |||| ||  |||
Db     135668 CAAAAGATATTAAATTTAGTGAAGATGCTCGTCGTGCCATGTTACGTGGTGTCGACCAAC 135727

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
              | ||||| || || |||||||| ||||| || ||||| |||||||| ||| | ||||
Db     135728 TAGCAAACGCAGTAAAAGTAACGCTTGGCCCAAAAGGTCGTAATGTTGTTTTAGAAAAAA 135787

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
               || |||||| ||  | |||||||| |||||||| |  |||||||||||| ||| | |
Db     135788 AATTCGGTTCTCCGTTAATTACAAATGATGGTGTAACGATTGCAAAAGAAATCGAATTAG 135847

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
              ||||    |||||||| |||||||  || | ||      ||||| ||| | ||||| |  |
Db     135848 AAGACCCATTTGAAAACATGGGCGCAAAACTTGTATCTGAAGTTGCTTCTAAAACCAATG 135907

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
              || ||||||| || |||||||||||| || || ||  | |||||||| ||  ||  ||||
Db     135908 ATGTTGCTGGGGACGGAACTACAACTGCTACCGTTTTAGCACAAGCAATGATTCAAGAAG 135967

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
              |  | |||   ||| |||||||    ||||| | | | ||   ||||| || || ||||
Db     135968 GCTTGAAAAACGTAACAGCTGGAGCAAATCCAGTAGGCGTTCGTCGTGGTATCGAAAAAG 136027

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
              | || |   ||| || | |||||  |||   |  ||  | || || | |     | || |
Db     136028 CCGTAGCAACTGCTATTGAAGAATTAAAAGCTATCTCTAAACCAATTGAAAGCAAAGAGT 136087

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
              || ||||||||||   ||||||||  |       |||| |  | | |||||  | | | |
Db     136088 CTATCGCTCAAGTTGCTGCCATTTCTTC---TGGTGATGAAGAAGTTGGTAAATTAATCG 136144

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
              | |||||  ||||     ||||| || |  |||||||| ||  |||| ||| | |||||
Db     136145 CAGAAGCAATGGAACGTGTTGGTAACGATGGTGTTATTACTATTGAAGAATCCAAAGGCT 136204

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
              ||| |||   |||||| || || |||||  ||| | ||||||||||||||     ||| |
Db     136205 TTGCAACAGAATTAGACGTAGTAGAAGGTATGCAATTTGACCGTGGCTACACTAGTCCTT 136264

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
              || | |||||| || | || |||||||     |  ||||| || || || ||   |  ||
Db     136265 ACATGGTAACTGATTCCGACAAAATGGAAGCAGTCCTTGAAAAACCATACATTTTAATTA 136324

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
              || ||||| || |  | |||  |||| ||  || || | ||||||||||||| | ||
Db     136325 CAGATAAAAAAATCAACAACATCCAAGAAATCTTATCAGTTTTAGAACAAGTTGTTCAAC 136384

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
              |   |||||| |  | ||||| || ||||| || ||||| |||||  |  |||||  || |
Db     136385 AAGGTCGTCCAATGTTAATTATCGCGGAAGATGTTGAAGGGGAAGCTCAAGCAACTCTAG 136444

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
              || | || || || || ||| || |  | || || || || ||||||||||||| ||||
Db     136445 TACTAAACAAACTTCGCGGAACATTTAACGTAGTTGCAGTGAAAGCTCCTGGTTTCGGTG 136504

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
              | |||||||||| ||||| |||||||||||  |  | |||||  ||| ||| |||    |
Db     136505 ACCGCCGTAAAGCAATGCTAGAAGATATTGCTGTTTTAACTGGTGGACAAGTAATCACAG 136564

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
              |||| |  ||| || | ||  |||  | ||  |  |  | ||||||||||  |
Db     136565 AAGACCTAGGTTTAGAACTGAAAACAGCAACAGTTGATCAACTTGGAACAGCTAACAAAG 136624

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
              ||||| |   |||||  |||| || ||||| || || || ||  | ||   | | || |
Db     136625 TAGTTGTAACAAAAGATGATACAACAATCGTAGAAGGAGCAGGCGATTCCACACAAATCA 136684

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
              ||||| || ||||| ||||| ||||| |||||||| |  || || |||| | ||
Db     136685 GCGCTCGCGTAAACCAAATCCGTGCGCAAATGGAAGAAACTACTTCTGAATTTGATAGAG 136744

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
              ||||| | |||||||| | |||||||||  |||||||||||||| |  | || || |
Db     136745 AAAAATTACAAGAACGTTTAGCAAAACTTGCAGGTGGAGTAGCTGTTGTTAAAGTCGGCG 136804

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
              |||| |||||||| ||  | |||||     ||    ||| | |||||||| || ||  | |
Db     136805 CTGCAACTGAAACAGAGCTAAAAGAACGTAAATTACGTATTGAAGATGCCCTTAACTCTA 136864

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
              |  | || ||  |||||||||||| ||  ||||||||||||||||  |||| |
Db     136865 CTCGTGCAGCAGTAGAAGAAGGTATCGTAGCTGGTGGTGGTACTGCCCTTGTAAGTATTT 136924

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
              ||    | | |  |  ||  || || |  |||| |   | ||  | || |||
Db     136925 ACAATAAAGTAGCAGCACTAGAA---GCAGAAGGCGATGTAGAAACAGGTATCAACATCG 136981

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
              | | |||||| || ||||| ||  | ||||| || ||    || |||||   ||||||||
Db     136982 TGCTTCGTTCCCTAGAAGAACCAGTTCGTCAGATCGCGCATAACGCTGGACTTGAAGGTT 137041

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
              || || | |||||    |     |   | || |||  |||| || || || || ||  |
Db     137042 CTGTTATCGTAGAACGCTTGAAACACGAAGCAGTTGGCGTTGGTTTCAACGCAGCTAATG 137101

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
              | ||||  | || | ||| | ||||||  |  ||||||| | |||||||| |||    |
Db     137102 GCGAATGGGTAAACATGATTGACGCTGGTATTGTTGATCCAACAAAAGTTACTCGTTCCG 137161

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
              |||||||||| ||  |||| || ||  |  | ||  ||||||| |||   | | | || |
Db     137162 CATTACAAAACGCTTCATCTGTTGCTGCTCTTCTATTAACTACCGAAGCTGTTGTAGCAG 137221

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627
              |  |||||||    ||     |    || |     || ||||| || ||||| |
Db     137222 ATCAACCAGATGAAAACGGCCCAGCAGCAGGCCCAGATATGGGAATGGGCGGCATGGGCG 137281

Qy       1628 GTATG 1632
              |||||
Db     137282 GTATG 137286


RESULT 8
ABQ69245/c
ID   ABQ69245 standard; DNA; 3011208 BP.
XX
AC   ABQ69245;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria innocua DNA sequence #684.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria innocua.
XX
PN   WO200228891-A2.
```

```
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators  -
XX
PS   Claim 5; SEQ ID 2058; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;

  Query Match             44.3%;  Score 729;  DB 24;  Length 3011208;
  Best Local Similarity   66.4%;  Pred. No. 3.3e-157;
  Matches 1079;  Conservative    0;  Mismatches  540;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db    2196334 CAAAAGATATTAAATTTAGTGAAGATGCTCGTCGTGCCATGTTACGTGGTGTCGACCAAC 2196275

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db    2196274 TAGCAAACGCAGTAAAAGTAACGCTTGGCCCAAAAGGTCGTAATGTTGTTTTAGAAAAAA 2196215

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db    2196214 AATTCGGTTCTCCGTTAATTACAAATGATGGTGTAACGATTGCAAAAGAAATCGAATTAG 2196155

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db    2196154 AAGACCCATTTGAAAACATGGGCGCAAAACTTGTATCTGAAGTTGCTTCTAAAACCAATG 2196095

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db    2196094 ATGTTGCTGGGGACGGAACTACAACTGCTACCGTTTTAGCACAAGCAATGATTCAAGAAG 2196035

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db    2196034 GCTTGAAAAACGTAACAGCTGGAGCAAATCCAGTAGGCGTTCGTCGTGGTATCGAAAAAG 2195975

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db    2195974 CCGTAGCAACTGCTATTGAAGAATTAAAAGCTATCTCTAAACCAATTGAAAGCAAAGAGT 2195915

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db    2195914 CTATCGCTCAAGTTGCTGCCATTTCTTC---TGGTGATGAAGAAGTTGGTAAATTAATCG 2195858

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db    2195857 CAGAAGCAATGGAACGTGTTGGTAACGATGGTGTTATTACTATTGAAGAATCCAAAGGCT 2195798

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db    2195797 TTGCAACAGAATTAGACGTAGTAGAAGGTATGCAATTTGACCGTGGCTACACTAGTCCTT 2195738

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db    2195737 ACATGGTAACTGATTCCGACAAAATGGAAGCAGTCCTTGAAAAACCATACATTTTAATTA 2195678

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db    2195677 CAGATAAAAAAATCAACAACATCCAAGAAATCTTATCAGTTTTAGAACAAGTTGTTCAAC 2195618

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db    2195617 AAGGTCGTCCAATGTTAATTATCGCGGAAGATGTTGAAGGGGAAGCTCAAGCAACTCTAG 2195558

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db    2195557 TACTAAACAAACTTCGCGGAACATTTAACGTAGTTGCAGTGAAAGCTCCTGGTTTCGGTG 2195498

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db    2195497 ACCGCCGTAAAGCAATGCTAGAAGATATTGCTGTTTTAACTGGTGGACAAGTAATCACAG 2195438

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db    2195437 AAGACCTAGGTTTAGAACTGAAAACAGCAACAGTTGATCAACTTGGAACAGCTAACAAAG 2195378

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db    2195377 TAGTTGTAACAAAAGATGATACAACAATCGTAGAAGGAGCAGGCGATTCCACACAAATCA 2195318

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db    2195317 GCGCTCGCGTAAACCAAATCCGTGCGCAAATGGAAGAAACTACTTCTGAATTTGATAGAG 2195258

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db    2195257 AAAAATTACAAGAACGTTTAGCAAAACTTGCAGGTGGAGTAGCTGTTGTTAAAGTCGGCG 2195198

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db    2195197 CTGCAACTGAAACAGAGCTAAAAGAACGTAAATTACGTATTGAAGATGCCCTTAACTCTA 2195138

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db    2195137 CTCGTGCAGCAGTAGAAGAAGGTATCGTAGCTGGTGGTGGTACTGCCCTTGTAAGTATTT 2195078

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db    2195077 ACAATAAAGTAGCAGCACTAGAA---GCAGAAGGCGATGTAGAAACAGGTATCAACATCG 2195021

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db    2195020 TGCTTCGTTCCCTAGAAGAACCAGTTCGTCAGATCGCGCATAACGCTGGACTTGAAGGTT 2194961

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db    2194960 CTGTTATCGTAGAACGCTTGAAACACGAAGCAGTTGGCGTTGGTTTCAACGCAGCTAATG 2194901

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db    2194900 GCGAATGGGTAAACATGATTGACGCTGGTATTGTTGATCCAACAAAAGTTACTCGTTCCG 2194841

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db    2194840 CATTACAAAACGCTTCATCTGTTGCTGCTCTTCTATTAACTACCGAAGCTGTTGTAGCAG 2194781

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627
Db    2194780 ATCAACCAGATGAAAACGGCCCAGCAGCAGGCCCAGATATGGGAATGGGCGGCATGGGCG 2194721
```

```
Qy       1628 GTATG 1632
              |||||
Db    2194720 GTATG 2194716


RESULT 9
ABA03041/c
ID   ABA03041 standard; DNA; 2944528 BP.
XX
AC   ABA03041;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes EGD-e genome sequence.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease; ds.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JA;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides -
XX
PS   Claim 1; SEQ ID No 1; 192pp; French.
XX
CC   The present sequence is the genome sequence of Listeria monocytogenes
CC   EGD-e. This sequence and fragments of this sequence are useful for
CC   selecting probes and primers for detecting genes in L. monocytogenes and
CC   related organisms, and to study genetic polymorphisms and other genomes.
CC   Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC   useful for raising specific antibodies, identification of L.
CC   monocytogenes and related organisms, and for biosynthesis and
CC   biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC   proteins encoded by it are also useful for selecting compounds that
CC   regulate gene expression and cell replication and modulate L.
CC   monocytogenes-related diseases. In addition, this sequence and proteins
CC   encoded by it are useful in pharmaceutical and vaccines compositions for
CC   the treatment or prevention of infections by L. monocytogenes and related
CC   organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

  Query Match             43.9%;  Score 722.6;  DB 24;  Length 2944528;
  Best Local Similarity   66.2%;  Pred. No. 9.9e-156;
  Matches 1075;  Conservative    0;  Mismatches  544;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
              | ||||| ||   ||| ||  | || |||       | | ||| ||||| || |||
Db    2149308 CAAAAGATATTAAATTTAGTGAAGATGCTCGTCGTGCCATGTTACGTGGTGTCGACCAAC 2149249

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
              | ||||| || || |||||||| ||||| || ||||| |||||||| ||| | || |||
Db    2149248 TAGCAAACGCAGTAAAAGTAACGCTTGGCCCAAAAGGTCGTAATGTTGTTTTAGAGAAGA 2149189

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
               || ||||| ||  | |||||||| ||||| ||| |  ||||||||||||||||| | |
Db    2149188 AATTCGGTTCTCCGTTAATTACAAATGATGGGGTAACAATTGCAAAAGAAATTGAATTAG 2149129

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
              ||||    |||||||||||||| ||  || | ||    ||||| ||| | ||||| | |
Db    2149128 AAGACCCATTTGAAAATATGGGAGCAAAACTTGTATCTGAAGTTGCTTCTAAAACCAATG 2149069

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
              || ||||||| || |||||||||||||| || ||  | || ||||| ||  || |||| |
Db    2149068 ATGTTGCTGGGGACGGAACTACAACAGCTACCGTTTTAGCGCAAGCAATGATTCAAGAAG 2149009

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
              | |||||  |||| |||||||    || ||  | | | ||   || || || || ||||
Db    2149008 GCTTAAAAAATGTAACAGCTGGAGCAAACCCAGTAGGCGTTCGCCGCGGTATCGAAAAAG 2148949

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
              | || |   | | ||   ||||| |||  |  ||| | || || | |     | || | |
Db    2148948 CCGTAGCAACAGCTATCGAAGAATTAAAAGCTATTTCTAAACCAATTGAAAGCAAAGAGT 2148889

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
              || ||||||||||    | ||||||  |      ||||  |  | | |||||  | || |
Db    2148888 CTATCGCTCAAGTTGCTGCTATTTCTTC---TGGTGATGAAGAAGTTGGTAAATTAATCG 2148832

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
              | |||||  ||||     |||||  || |  || |||||||  ||||  |||  | ||||||
Db    2148831 CAGAAGCAATGGAACGCGTTGGTAACGACGGCGTTATCACTATTGAAGAATCCAAAGGTT 2148772

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
              |||  |||   |||||| || || |||||  ||| | || |||||  ||||||      || |
Db    2148771 TTGCAACAGAATTAGACGTAGTAGAAGGTATGCAATTCGACCGCGGCTACACTAGCCCTT 2148712

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
              || | ||||||  || | || |||||||     ||  |||||  || || || ||  |  ||
Db    2148711 ACATGGTAACTGATTCCGATAAAATGGAAGCTGTTCTTGAAAAACCATACATTTTAATTA 2148652

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
              || ||||| || |  |  ||  |||| ||  |||||  | ||||||||||| | | ||
Db    2148651 CAGACAAAAAAATCAACAATATTCAAGAAATCTTACCAGTTTTAGAACAAGTCGTTCAAC 2148592

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
              |   |||||| |  | || ||||| ||||| || ||||| ||||| |  ||||| ||||
Db    2148591 AAGGTCGTCCAATGTTAATCATTGCGGAAGATGTTGAAGGGGAAGCTCAAGCAACTCTTG 2148532

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
              || | || || || || ||| || |  | || |||||||| |||||||||||||| ||||
Db    2148531 TACTAAACAAACTTCGCGGAACATTTAACGTAGTAGCCGTGAAAGCTCCTGGTTTCGGTG 2148472

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
              | || |||||||| ||||| |||||||||||||  | || || ||| ||| ||    |
Db    2148471 ATCGTCGTAAAGCAATGCTAGAAGATATTGCTATTTTAACAGGCGGACAAGTCATTACAG 2148412

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
              |||| |  ||  || | ||  |||  | |   |  |  | ||||||||||||    |
Db    2148411 AAGACCTAGGATTAGAACTAAAAACAGCTACTGTTGATCAACTTGGAACAGCTAACAAAG 2148352

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
              |||| |   |||||  |||| || |||| || || || ||  | ||  | | ||||
Db    2148351 TAGTCGTAACAAAAGATGATACAACAATCGTAGAAGGAGCAGGCGATTCCACACAAATTA 2148292

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
              |||||  || || ||||| |||||  |||||||| |  || || | |||| | |
Db    2148291 GCGCTCGCGTAAACCAAATCCGTGCGCAAATGGAAGAAACTACTTCTGAATTTGATAGAG 2148232

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
              ||||| | |||||||| | ||||||||||  |||||||||||||| |  | || || |
```

```
Db    2148231 AAAAATTACAAGAACGGTTAGCAAAACTTGCAGGTGGAGTAGCTGTTGTTAAAGTCGGCG 2148172

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207

Db    2148171 CTGCAACTGAAACAGAGCTAAAAGAACGTAAATTACGTATTGAAGATGCACTTAACTCTA 2148112

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267

Db    2148111 CTCGCGCAGCTGTAGAAGAAGGTATCGTAGCTGGTGGTGGTACTGCCCTTGTAAGTATTT 2148052

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327

Db    2148051 ACAATAAAGTAGCAGCACTTGAA---GCAGAAGGCGACGTAGAAACAGGTATCAACATTG 2147995

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387

Db    2147994 TGCTTCGTTCTCTAGAAGAACCAGTTCGCCAAATCGCACATAACGCTGGACTTGAAGGTT 2147935

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447

Db    2147934 CGGTTATCGTTGAACGCTTGAAACACGAAGCAGTTGGTGTTGGTTTCAATGCAGCAAACG 2147875

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507

Db    2147874 GCGAATGGGTAAACATGATTGATGCAGGTATTGTGGATCCAACTAAAGTAACACGTTCCG 2147815

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567

Db    2147814 CTCTACAAAATGCATCATCTGTTGCAGCGCTTCTATTAACTACAGAAGCAGTCGTAGCAG 2147755

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGTG 1627

Db    2147754 ACAAACCAGACGAAAACGGCCCAGCAGCTGTTCCTGATATGGGAATGGGTGGCATGGGCG 2147695

Qy       1628 GTATG 1632

Db    2147694 GCATG 2147690
```

RESULT 10
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H.influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFS can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 43.8%; Score 721; DB 17; Length 1830121;
Best Local Similarity 65.7%; Pred. No. 2e-155;
Matches 1083; Conservative 0; Mismatches 560; Indels 6; Gaps 2;

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db     565350 ATGGCAGCAAAAGACGTAAAATTTGGTAATGACGCACGCGTAAAAATGCTTAAAGGCGTG 565409

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Db     565410 AATGTATTAGCGGATGCAGTAAAAGTAACCCTTGGCCCGAAAGGTCGTCATGTAATTTTA 565469

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180

Db     565470 GATAAATCATTTGGCGCACCAACTATCACTAAAGACGGTGTGTCTGTTGCTCGTGAAATC 565529

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240

Db     565530 GAATTAGAAGATAAATTCGAAAACATGGGGGCACAAATGGTGAAAGAAGTGGCATCTAAA 565589

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Db     565590 GCAAATGATGCTGCAGGTGACGGTACAACTACTGCAACTGTACTTGCGCAAGCTATCGTA 565649

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Db     565650 AATGAAGGCTTGAAAGCAGTAGCTGCAGGTATGAATCCAATGGATTTAAAACGTGGTATT 565709

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db     565710 GATAAAGCAGTAAGTGCGGTTGTTTCTGAACTTAAAAATTTATCTAAACCTTGTGAAACA 565769

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db     565770 GCAAAAGAAATTGAACAAGTAGGGACTATTTCTGCAAACTCTGACAGCATTGTGGGTCAA 565829

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db     565830 TTAATCTCACAAGCAATGGAAAAAGTGGGCAAGGAAGGTGTAATTACTGTTGAAGATGGT 565889

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db     565890 ACGGGTCTTGAAGATGAATTAGATGTGGTTGAAGGCATGCAATTAGATAATGGTTACCTT 565949

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db     565950 TCTCCATATTTCATCAACAAACCAGAAACTGCAACCGTTGAATTAGATAATCCATATCTT 566009

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db     566010 CTTCTTGTAGATAAAAAAATCTCTAACATTCGTGAATTACTTCCCGTGTTAGAAGGCGTT 566069

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
```

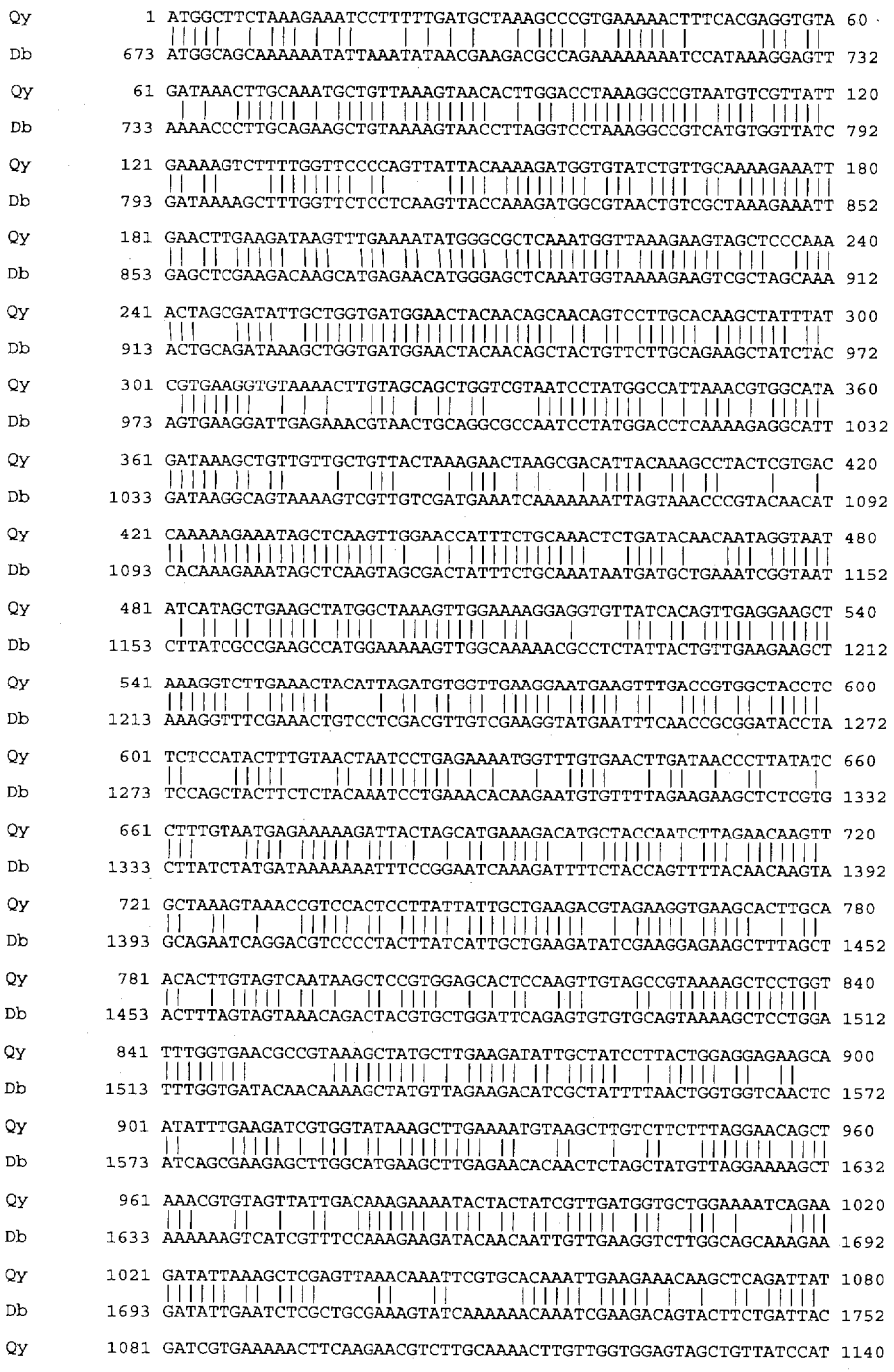
```
Db     566070 GCGAAAGCAGGTAAACCGTTATTAATCATCGCTGAAGACGTGGAAGGCGAAGCGCTTGCA 566129

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db     566130 ACCTTAGTGGTAAACACTATGCGCGGTATCGTGAAAGTTGCAGCCGTGAAAGCACCAGGT 566189

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db     566190 TTTGGTGATCGTCGTAAAGCGATGTTACAAGATATTGCAATTTTAACAGCGGGTACAGTG 566249

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db     566250 ATTTCTGAAGAAATTGGTATGGAGCTTGAAAAAGCAACATTGGAAGATTTAGGTCAAGCA 566309

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db     566310 AAACGTGTTGTTATCAATAAAGATAACACAACCATTATTGATGGTATCGGCGATGAAGCA 566369

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db     566370 CAAATCAAAGGTCGTGTGGCTCAAATTCGTCAGCAAATCGAAGAATCAACTTCTGACTAC 566429

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db     566430 GACAAAGAAAAACTCCAAGAACGCGTGGCTAAATTAGCAGGCGGTGTGGCAGTAATCAAA 566489

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db     566490 GTGGGCGCAGCAACTGAAGTTGAAATGAAAGAGAAAAAAGATCGTGTAGATGATGCATTA 566549

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db     566550 CATGCAACTCGTGCAGCGGTTGAAGAAGGTATCGTTGCTGGTGGTGGCGTTGCATTAGTT 566609

Qy       1261 CGCTC---CATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317

Db     566610 CGCGCASSAGCGAAAGTTGCAGCAAGCCTAAAAGGTGACAACGAAGAACAAAATGTGGGT 566669

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377

Db     566670 ATCAAACTTGCATTACGTGCTATGGAAGCGCCTTTACGTCAAATCGTCACTAACGCAGGT 566729

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437

Db     566730 GAAGAAGCTTCAGTTGTGGCAAGTGCGGTTAAAAATGGCGAAGGAAACTTTGGTTATAAT 566789

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497

Db     566790 GCAGGAACAGAACAGTACGGCGATATGATCGAAATGGGTATCTTAGATCCAACTAAAGTC 566849

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557

Db     566850 ACTCGTTCTGCATTGCAATTCGCAGCTTCTGTAGCAGGCTTAATGATTACCACTGAATGT 566909

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617

Db     566910 ATGGTAACTGATCTTCCAAAAGATGATAAAG---CCGATTTAGGTGCTGCTGGAATGGGC 566966

Qy       1618 GGTATGGGTGGTATGGACGGTATGTACTA 1646

Db     566967 GGTATGGGTGGAATGGGCGGAATGATGTA 566995
```

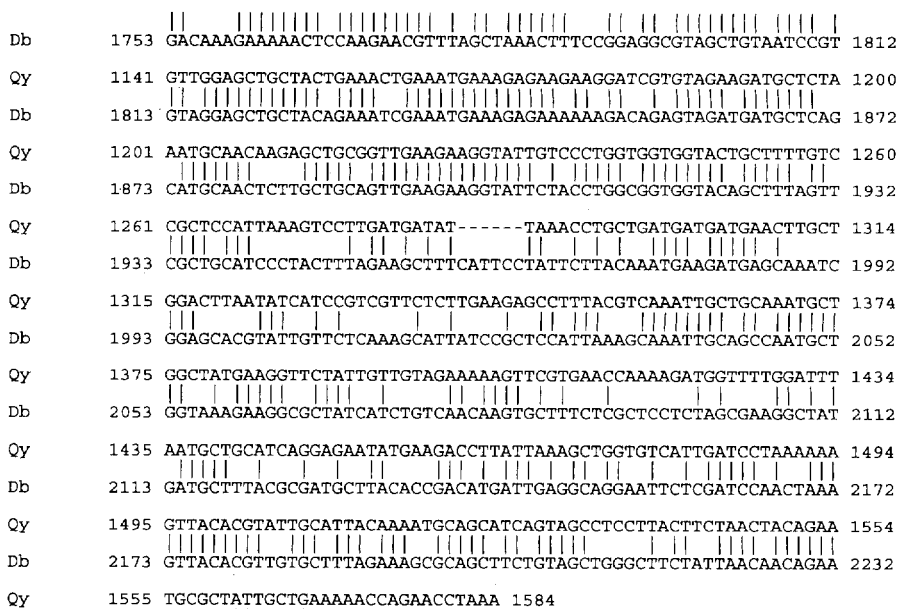
RESULT 11
AAQ13136
ID AAQ13136 standard; DNA; 2465 BP.
XX
AC AAQ13136;
XX
DT 25-MAR-2003 (updated)
DT 17-DEC-2001 (updated)
DT 22-OCT-1991 (first entry)
XX
DE Hyp operon.
XX
KW Antibodies; heat shock; hypersensitive; allergen; HSP60; GroEL;
KW GroES; ss.
XX
OS Chlamydia psittaci GPIC.
XX

FH	Key	Location/Qualifiers
FT	CDS	314..622
FT		/*tag= a
FT		/label= hypA
FT	CDS	673..2307
FT		/*tag= b
FT		/label= hypB
FT	-35_signal	83..88
FT		/*tag= c
FT	-10_signal	120..127
FT		/*tag= d
FT	RBS	301..304
FT		/*tag= e
FT	RBS	656..659
FT		/*tag= f
FT	terminator	2404..2440
FT		/*tag= g
FT		/note=" dyad symmetry; rho-independent"
FT	misc_feature	350..385
FT		/*tag= h
FT		/label= oligo
FT		/note= "used for Northern blot analysis"
FT	misc_feature	2152..2187
FT		/*tag= i
FT		/label= oligo
FT		/note= "used in Northern blot analysis"

XX
PN USN7531317-N.
XX
PD 09-JUL-1991.
XX
PF 31-MAY-1990; 90US-0143560.
XX
PR 31-MAY-1990; 90US-0143560.
XX
PA (USSH) NAT INST OF HEALTH.
XX
DR WPI; 1991-245693/33.
DR P-PSDB; AAR13334, AAR13335.
XX
PT DNA encoding HypA and HypB Chlamydia proteins - used to develop
PT prods. for detection of and vaccines against Chlamydia infection.
XX
PS Disclosure; Fig 5; 51pp; English.
XX
CC The sequence was obtd. from clone pGP57 isolated from a DNA library
CC prepd. from C. psittaci genomic DNA. The two ORFs encode HypA and
CC HypB hypersensitivity proteins, regulated by a heat shock type
CC promoter region. The DNA can be used to produce recombinant
CC proteins or to design probes for the detection of Chlamydial
CC infection. The Hyp proteins can be used to to raise antibodies and
CC in vaccines. The Hyp B protein elicits a cell-mediated immune
CC response so can be used as a skin test antigen.
CC See also AAQ13137.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2465 BP; 853 A; 467 C; 509 G; 636 T; 0 other;

Query Match 43.3%; Score 713.6; DB 12; Length 2465;
Best Local Similarity 66.0%; Pred. No. 1.3e-154;
Matches 1050; Conservative 0; Mismatches 534; Indels 6; Gaps 1;



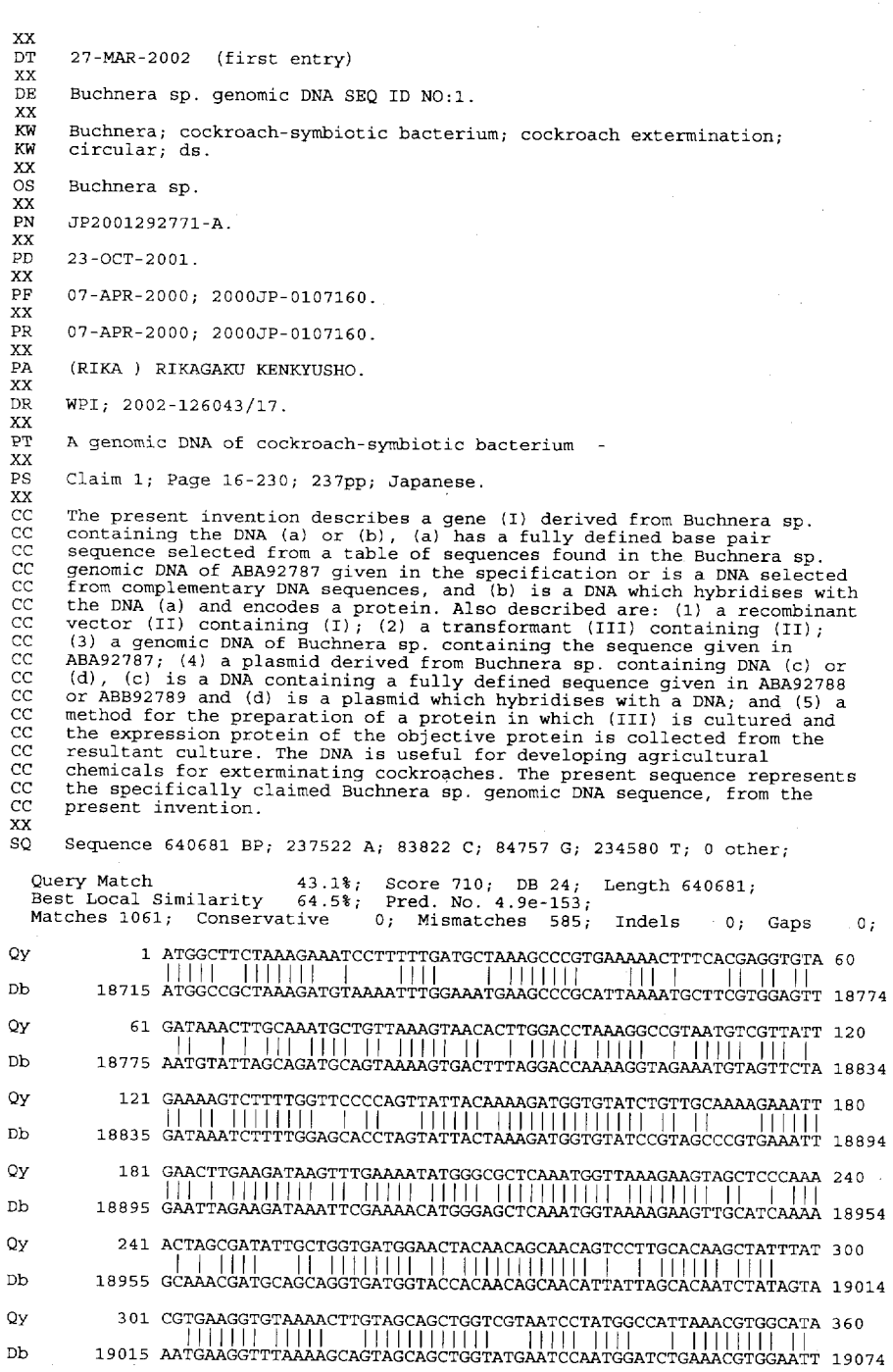


```
RESULT 12
AAQ13137
ID   AAQ13137 standard; DNA; 2223 BP.
XX
AC   AAQ13137;
XX
DT   25-MAR-2003  (updated)
DT   17-DEC-2001  (updated)
DT   22-OCT-1991  (first entry)
XX
DE   Hyp operon.
XX
KW   Antibodies; heat shock; hypersensitive; allergen; HSP60; GroEL; ss.
XX
OS   Chlamydia trachomatis serovar A.
XX
FH   Key             Location/Qualifiers
FT   CDS             35..343
FT                   /*tag=  a
FT                   /label= hypA
FT   CDS             381..2015
FT                   /*tag=  b
FT                   /label= hypB
FT   RBS             23..26
FT                   /*tag=  c
FT   RBS             35..38
FT                   /*tag=  d
FT   terminator      2046..2072
FT                   /*tag=  e
FT                   /note=" dyad symmetry; rho-independent"
XX
PN   USN7531317-N.
XX
PD   09-JUL-1991.
XX
PF   31-MAY-1990;   90US-0143560.
XX
```

```
PR   31-MAY-1990;   90US-0143560.
XX
PA   (USSH ) NAT INST OF HEALTH.
XX
DR   WPI; 1991-245693/33.
DR   P-PSDB; AAR13336, AAR13337.
XX
PT   DNA encoding HypA and HypB Chlamydia proteins - used to develop
PT   prods. for detection of and vaccines against Chlamydia infection.
XX
PS   Disclosure; Fig 7; 51pp; English.
XX
CC   The sequence was obtd. from clone pTA571 isolated from a DNA library
CC   prepd. from C. trachomatis genomic DNA.  The two ORFs encode HypA
CC   and HypB hypersensitivity proteins.  The DNA can be used to produce
CC   recombinant proteins or to design probes for the detection of
CC   Chlamydial infection.  The Hyp proteins can be used to to raise
CC   antibodies and in vaccines.  The Hyp B protein elicits a cell-
CC   mediated immune response so can be used as a skin test antigen.
CC   See also AAQ13136.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government-owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 2223 BP; 745 A; 408 C; 489 G; 581 T; 0 other;

  Query Match             43.1%;  Score 710.4;  DB 12;  Length 2223;
  Best Local Similarity   65.9%;  Pred. No. 6.9e-154;
  Matches 1048;  Conservative    0;  Mismatches  536;  Indels    6;  Gaps    1;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db        381 ATGGTCGCTAAAAACATTAAATACAACGAAGAAGCCAGAAAGAAAATTCAAAAAGGAGTT 440

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        441 AAGACTTTAGCTGAAGCTGTAAAAGTCACTCTAGGGCCTAAAGGACGACATGTTGTCATA 500

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        501 GATAAAAGCTTCGGATCCCCTCAAGTAACTAAAGATGGTGTTACCGTTGCGAAAGAAGTT 560

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        561 GAGCTTGCCGACAAACATGAAAATATGGGCGCTCAAATGGTCAAAGAAGTCGCCAGCAAA 620

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        621 ACTGCTGACAAAGCTGGAGACGGAACTACAACAGCTACTGTTCTTGCTGAAGCTATCTAT 680

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        681 ACAGAAGGATTACGCAATGTAACAGCTGGAGCAAATCCAATGGACCTCAAACGAGGTATT 740

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        741 GATAAAGCTGTTAAGGTTGTTGTTGATCAAATCAAAAAAATCAGCAAGCCTGTTCAGCAT 800

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        801 CATAAAGAAATTGCTCAAGTTGCAACAATTTCTGCTAATAATGATGCAGAAATCGGGAAT 860

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        861 CTGATTGCTGAAGCAATGGAGAAAGTTGGTAAAAACGGCTCTATCACTGTTGAAGAAGCA 920

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        921 AAAGGATTTGAAACCGTTTTGGATGTTGTTGAAGGAATGAATTTCAATAGAGGTTACCTC 980

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        981 TCTAGCTACTTCGCAACAAATCCAGAAACTCAAGAATGTGTATTAGAAGACGCTTTGGTT 1040

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       1041 CTAATCTACGATAAGAAAATTTCTGGGATCAAAGATTTCCTTCCTATTTTACAACAAGTT 1100

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       1101 GCTGAATCCGGCCGTCCTCTTCTTATTATAGCAGAAGACATTGAAGGCGAAGCTTTAGCT 1160

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       1161 ACTTTGGTCGTGAACAGAATTCGTGGAGGATTCCGGGTTTGCGCAGTTAAAGCTCCAGGC 1220

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       1221 TTTGGAGATAGAAGAAAAGCTATGTTGGAAGACATCGCTATCTTAACTGGCGGTCAACTC 1280

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       1281 ATTAGCGAAGAGTTGGGCATGAAATTAGAAAACGCTAACTTAGCTATGTTAGGTAAAGCT 1340

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       1341 AAAAAAGTTATCGTTTCTAAAGAAGACACGACCATCGTCGAAGGAATGGGTGAAAAAGAA 1400

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1401 GCTTTAGAAGCTCGTTGCGAAAGCATCAAAAAACAAATTGAAGACAGCTCTTCTGATTAC 1460

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1461 GATAAAGAAAAACTCCAAGAGCGTCTTGCTAAGCTCTCTGGTGGAGTAGCAGTCATTCGC 1520

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1521 GTTGGAGCTGCAACAGAGATTGAGATGAAAGAGAAAAAAGATCGTGTAGACGATGCTCAA 1580

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1581 CATGCTACAATCGCTGCTGTTGAAGAAGGAATTCTTCCTGGTGGAGGAACAGCATTAATC 1640

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTG------CTGATGATGATGAACTTGCT 1314
Db       1641 CGTTGTATCCCTACTCTTGAAGCCTTCTTGCCAATGTTGACTAATGAAGATGAGCAAATT 1700

Qy       1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCT 1374
Db       1701 GGAGCTCGCATTGTTTTGAAAGCTCTTTCCGCTCCTTTGAAACAAATTGCTGCAAACGCA 1760

Qy       1375 GGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTT 1434
Db       1761 GGAAAAGAAGGTGCTATCATCTTCCAACAAGTTATGTCCCGTTCTGCGAACGAAGGATAT 1820

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db       1821 GATGCATTGCGTGATGCATACACAGATATGCTTGAAGCTGGTATTTTAGATCCTGCTAAA 1880

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db       1881 GTAACCCGTTCTGCTTTAGAAAGCGCGGCTTCCGTAGCTGGATTACTTTTGACAACAGAA 1940

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAA 1584
Db       1941 GCTCTCATTGCAGAGATTCCAGAAGAAAAA 1970

RESULT 13
ABA92787
ID   ABA92787 standard; DNA; 640681 BP.
XX
AC   ABA92787;
```

XX
DT 27-MAR-2002 (first entry)
XX
DE Buchnera sp. genomic DNA SEQ ID NO:1.
XX
KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
XX
OS Buchnera sp.
XX
PN JP2001292771-A.
XX
PD 23-OCT-2001.
XX
PF 07-APR-2000; 2000JP-0107160.
XX
PR 07-APR-2000; 2000JP-0107160.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
DR WPI; 2002-126043/17.
XX
PT A genomic DNA of cockroach-symbiotic bacterium -
XX
PS Claim 1; Page 16-230; 237pp; Japanese.
XX
CC The present invention describes a gene (I) derived from Buchnera sp.
CC containing the DNA (a) or (b), (a) has a fully defined base pair
CC sequence selected from a table of sequences found in the Buchnera sp.
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (I); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC (d), (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABB92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and
CC the expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents
CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC present invention.
XX
SQ Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;

Query Match 43.1%; Score 710; DB 24; Length 640681;
Best Local Similarity 64.5%; Pred. No. 4.9e-153;
Matches 1061; Conservative 0; Mismatches 585; Indels 0; Gaps 0;



```
Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db      20155 GCTACTGATGAATATGGTGACATGATAGATTTTGGTATATTAGATCCAACTAAAGTTACA 20214

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db      20215 CGTTCTGCTTTACAGTATGCTGCTTCTGTCGCTGGTCTAATGATCACAACAGAATGTATG 20274

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db      20275 GTAACTGACTTGCCTAAAGAAGATAAATCTTCTGATTCTAGTTCTTCTCCAGCAGGAGGA 20334

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
Db      20335 ATGGGTGGCATGGGTGGAATGATGTA 20360


RESULT 14
ABA90521
ID   ABA90521 standard; DNA; 2365589 BP.
XX
AC   ABA90521;
XX
DT   16-MAY-2002  (first entry)
XX
DE   Genomic sequence of Lactococcus lactis IL1403.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX
OS   Lactococcus lactis IL1403.
XX
PN   FR2807446-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-0004630.
XX
PR   11-APR-2000; 2000FR-0004630.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,  Sorokine A,  Renault P,  Ehrlich SD;
XX
DR   WPI; 2002-043418/06.
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species -
XX
PS   Claim 1; SEQ ID 1; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC   nucleic acid sequence is useful in the detection and/or amplification of
CC   nucleic acid sequence, particularly to identify Lactococcus lactis or
CC   related species. The proteins of the invention are useful for the
CC   biosynthesis or biodegradation of a composition of interest. The
CC   invention helps research in lactic bacteria, particularly useful in the
CC   production of yogurt and cheese.
CC   Note: The sequence data for this patent is based on equivalent patent
CC   WO200177334 (published 18-OCT-2001) which is available in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

  Query Match           43.0%;  Score 708.4;  DB 24;  Length 2365589;
  Best Local Similarity   66.5%;  Pred. No. 1.7e-152;
  Matches 1047;  Conservative    0;  Mismatches  521;  Indels    6;  Gaps    2;

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db     398900 TCAAAAGATATTAAATTTTCAAGTGATGCTAGAACAGCGATGATGCGTGGGATTGATATT 398959

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db     398960 CTTGCTGATACAGTAAAAACAACCCTTGGACCAAAAGGTCGCAATGTTGTTCTTGAAAAA 399019

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db     399020 TCATATGGATCACCTTTAATTACAAACGATGGTGTAACGATTGCCAAAGAGATTGAACTT 399079

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db     399080 GAAGATCATTTTGAAAATATGGGAGCTAAACTTGTTTCAGAAGTTGCTTCAAAAACAAAT 399139

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db     399140 GATATCGCAGGTGACGGTACAACAACAGCGACAGTTTTGACACAAGCTATCGTTCGTGAA 399199

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db     399200 GGTTTAAAAAATGTTACCGCAGGTGCAAATCCAGTTGGTATTCGCCGAGGAATTGAACTT 399259

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db     399260 GCTGCTGAAACAGCTGTTGCGTCAATTAAAGAGATGGCAATTCCTGTCCACGATAAATCA 399319

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db     399320 GCAATTGCGCAAGTAGCTACCGTTTCATCACGTAGTGA---AAAAGTGGGTGAGTATATT 399376

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db     399377 TCTGATGCAATGGAACGTGTAGGTTCTGACGGAGTTATCACCATTGAAGAATCAAAAGGG 399436

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db     399437 ATGCAAACTGAACTCGATGTTGTTGAAGGAATGCAATTTGACCGTGGATATTTGAGTCAA 399496

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db     399497 TATATGGTTTCTAATACAGAAAAAATGGTTGCTGAATTAGATAATCCTTATATTCTTATT 399556

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db     399557 ACCGACAAAAAAATCTCAAACATTCAAGAAATTTTACCGCTCCTTGAACAAATCTTGAAA 399616

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db     399617 ACAAATCGTCCACTTCTTATTGTAGCTGATGATGTTGATGGAGAAGCATTGCCAACGCTT 399676

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db     399677 GTTCTTAATAAGATTAAAGGTGTCTTCAATGTTGTAGCTGTAAAAGCACCAGGATTTGGT 399736

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db     399737 GATCGTCGTAAAGCTCAATTGGAAGATTTGGCTATCTTGACAGGTGGTACAGTAATTACT 399796

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db     399797 GAAGAACTTGGTCTTGATCTTAAAGATGCTACTCTTGAAGCTTTAGGACAAGCAGCTAAG 399856

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db     399857 GCAACTGTGGATAAAGACCACACAACAATTGTTGAAGGTGCTGGTTCTGCTGATGCTATT 399916

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db     399917 TCTGATCGTGTAGCTATTATTAAAGCGCAAATTGAAAAAACAACTTCTGATTTTGATCGT 399976

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db     399977 GAAAAATTACAAGAACGTCTTGCAAAATTAGCTGGTGGTGTTGCAGTTGTTAAAGTTGGT 400036

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
```

```
Db      400037 GCTGCAACTGAAACAGAGCTCAAAGCAATGAAACTCTTGATTGAAGATGCACTTAATGCA 400096

Qy        1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266

Db      400097 ACACGTGCAGCCGTTGAGGAAGGTATTGTTTCTGGTGGTGGTACAGCACTTGTTAATGCT 400156

Qy        1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326

Db      400157 ATTGCTGCTTTGGATAAACTT---TCAGAAGAAGGAGATATTCAAACAGGTATTAACATT 400213

Qy        1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386

Db      400214 GTTCGCCGTGCTCTTGAAGAACCAGTTCGTCAAATTGCTGCCAATGCAGGATATGAAGGT 400273

Qy        1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446

Db      400274 TCAGTTATCATTGATAAACTTCGTTCAGAAGAAGTAGGTACAGGATTCAATGCTGCAACT 400333

Qy        1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506

Db      400334 GGTCAATGGGTTAATATGATTGAAGAAGGAATTGTTGACCCTGCAAAAGTAACTCGTTCA 400393

Qy        1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566

Db      400394 GCTTTGCAAAACGCGGCTTCTGTTGCTGGTTTAATTTTGACTACTGAAGCAGTAGTTGCT 400453

Qy        1567 GAAAAACCAGAACC 1580

Db      400454 AATAAGCCAGAACC 400467
```

RESULT 15
AAX86155
ID AAX86155 standard; DNA; 1661 BP.
XX
AC AAX86155;
XX
DT 22-SEP-1999 (first entry)
XX
DE DNA encoding a Streptococcus pyogenes heat shock protein (Hsp)60-2.
XX
KW Heat shock protein; Hsp60-2; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy; ss.
XX
OS Streptococcus pyogenes.
XX
PN WO9935270-A1.
XX
PD 15-JUL-1999.
XX
PF 29-DEC-1998; 98WO-CA01203.
XX
PR 31-DEC-1997; 97US-0001737.
XX
PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI Mizzen L, Wisniewski J;
XX
DR WPI; 1999-430397/36.
DR P-PSDB; AAY23904.
XX
PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
PS Claim 3; Fig 4A-B; 176pp; English.
XX
CC The present sequence encodes a heat shock protein, designated Hsp60-2.
CC The protein, its fragments, variants and fusion proteins, are
CC used to elicit or enhance an immune response against Streptococcus,
CC and to elicit a similar response to a target antigen fused to the
CC protein. Unlike other immunological carriers, Hsp60 proteins are not
CC immunosuppressive so provide an increased response to any conjugated or
CC fused antigen. Also, where used for cancer control, they lack the side
CC effects associated with endotoxins. They can also be used to detect
CC specific antibodies and in treatment or prevention of tumours
CC (e.g. sarcoma or cancers of breast, ovary, prostate, lung, pancreas or
CC liver). The Hsp60 polynucleotide is used for recombinant production
CC of the protein, as a source of primers and probes for detecting
CC streptococci in standard hybridization/amplification assays, and
CC therapeutically in gene therapy vectors.
XX
SQ Sequence 1661 BP; 512 A; 308 C; 407 G; 434 T; 0 other;

Query Match 42.8%; Score 704.2; DB 20; Length 1661;
Best Local Similarity 66.4%; Pred. No. 1.7e-152;
Matches 1044; Conservative 0; Mismatches 523; Indels 6; Gaps 2;

```
Qy           8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67

Db          19 CAAAAGAAATCAAATTTTCAGCAGATGCGCGTGCTGCCATGGTGCGCGGAGTTGATATGT 78

Qy          68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127

Db          79 TAGCAGATACCGTCAAAGTAACGCTTGGTCCTAAAGGGCGCAATGTTGTTCTTGAAAAAG 138

Qy         128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187

Db         139 CTTTTGGTTCTCCCTTAATTACTAATGACGGGGTAACCATTGCTAAAGAGATCGAATTAG 198

Qy         188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247

Db         199 AAGATCATTTTGAAAACATGGGAGCAAAATTGGTGTCTGAAGTGGCTTCTAAAACCAATG 258

Qy         248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307

Db         259 ATATTGCTGGTGATGGGACGACTACTGCAACAGTTTTGACACAAGCCATTGTTCATGAAG 318

Qy         308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367

Db         319 GACTAAAAAATGTGACAGCAGGTGCTAATCCAATTGGTATCCGTCGAGGCATTGAAACAG 378

Qy         368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427

Db         379 CAACAGCAACAGCTGTTGAAGCCTTGAAAGCCATTGCTCAACCTGTATCTGGCAAGGAAG 438

Qy         428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487

Db         439 CTATTGCTCAGGTCGCTGCAGTATCATCACGCTCTGA---AAAAGTTGGAGAGTATATCT 495

Qy         488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547

Db         496 CAGAAGCTATGGAGCGTGTGGGCAACGATGGTGTGATTACCATCGAAGAATCTCGAGGTA 555

Qy         548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607

Db         556 TGGAAACAGAACTTGAAGTGGTTGAAGGCATGCAATTTGACCGTGGTTACCTGTCTCAAT 615

Qy         608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667

Db         616 ACATGGTCACAGACAATGAAAAAATGGTTGCAGACCTTGAAAACCCATTTATCTTAATCA 675

Qy         668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727

Db         676 CGGATAAAAAAGTGTCAAACATCCAAGACATTTTGCCACTACTTGAGGAAGTTCTTAAAA 735

Qy         728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787

Db         736 CCAACCGTCCATTACTCATTATTGCAGATGATGTGGATGGTGAAGCACTTCCAACCCTTG 795

Qy         788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847

Db         796 TCTTGAACAAGATTCGTGGTACTTTCAATGTGGTTGCTGTCAAAGCGCCAGGATTTGGTG 855

Qy         848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
```

```
Db        856 ATCGTCGTAAAGCTATGCTTGAAGACATTGCTATCTTGACAGGTGGTACAGTGATTACAG 915

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967

Db        916 AGGATCTAGGACTTGAATTAAAAGATGCTACAATGACAGCCCTTGGACAGGCTGCTAAGA 975

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027

Db        976 TTACAGTTGATAAAGATAGCACAGTAATTGTTGAAGGTTCAGGAAGTTCAGAAGCTATTG 1035

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087

Db       1036 CTAACCGTATTGCACTGATTAAATCGCAATTAGAAACAACAACTTCTGACTTTGACCGTG 1095

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147

Db       1096 AAAAACTACAAGAACGTTTGGCGAAATTAGCTGGTGGTGTAGCTGTTATCAAAGTAGGAG 1155

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207

Db       1156 CTCCAACAGAGACAGCTTTAAAAGAAATGAAACTTCGCATTGAGGATGCTCTAAATGCTA 1215

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267

Db       1216 CACGTGCAGCCGTTGAAGAAGGTATCGTTGCTGGTGGTGGAACAGCACTTATTACGGTTA 1275

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327

Db       1276 TTGAAAAAGTAGCAGCTCTTGAGCTTGAGGGCGATGATG---CTACTGGACGTAACATTG 1332

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387

Db       1333 TGCTTCGTGCTCTAGAAGAGCCTGTACGTCAAATTGCTTTAAATGCTGGGTACGAAGGCT 1392

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447

Db       1393 CCGTAGTTATTGACAAGTTGAAAAACAGCCCTGCAGGAACAGGATTTAATGCTGCAACAG 1452

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507

Db       1453 GTGAGTGGGTTGATATGATTAAAACAGGAATCATTGACCCTGTCAAAGTAACACGATCAG 1512

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567

Db       1513 CGCTTCAAAATGCAGCTTCTGTAGCTAGTCTTATTTTGACAACAGAAGCAGTTGTTGCTA 1572

Qy       1568 AAAAACCAGAACC 1580

Db       1573 ATAAACCTGAACC 1585
```

Search completed: January 29, 2004, 13:35:33
Job time : 372 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:22:11 ; Search time 31.5 Seconds
(without alignments)
10056.496 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125847_11632/app_query.fasta_1.1799
-DB=PIR_76 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574_@CGN_1_1_43_@runat_28012004_125847_11632 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	62.9	551	2	JC2562	chaperonin groELx
2	1864	62.8	545	2	C95311	groEL2 chaperonin
3	1863	62.8	545	2	JN0509	heat shock protein
4	1862.5	62.8	547	2	F82783	60kDa chaperonin X
5	1852.5	62.4	546	2	B47073	chaperonin GroEL -
6	1845	62.2	550	2	A41468	60K heat shock pro
7	1844.5	62.1	546	2	S65596	heat shock protein
8	1843.5	62.1	547	2	I40331	Cpn60 protein (Gro
9	1843.5	62.1	552	2	S39765	chaperonin 60 - Co
10	1841	62.0	542	2	F95967	probable heat shoc
11	1840.5	62.0	544	2	AD2660	60 KDA chaperonin
12	1840.5	62.0	544	2	B97442	60K chaperonin (pr
13	1828.5	61.6	544	2	S23918	groEL protein - Ag
14	1825.5	61.5	547	2	B83098	GroEL protein PA43
15	1823	61.4	546	2	S22347	groEL - Brucella a
16	1823	61.4	550	2	S35309	heat shock protein
17	1822	61.4	546	2	I40342	heat shock protein
18	1817.5	61.2	544	2	B36917	heat shock protein
19	1816.5	61.2	541	2	JN0512	heat shock protein
20	1814.5	61.1	544	2	H81964	chaperonin 60kD su
21	1814.5	61.1	549	2	JC2564	heat shock protein
22	1814	61.1	546	2	AG3640	60K chaperonin gro
23	1812.5	61.1	545	2	S47530	chaperonin groEL -
24	1810.5	61.0	544	2	S37039	groEL protein - Ba
25	1808.5	60.9	544	2	C81021	chaperonin, 60 kDa
26	1805.5	60.8	547	2	B43606	heat shock protein
27	1802.5	60.7	544	2	S61300	heat shock protein
28	1799	60.6	546	2	B54539	heat shock protein
29	1794.5	60.5	544	2	S61302	heat shock protein
30	1794	60.4	546	2	S35311	heat shock cognate
31	1791.5	60.4	547	2	B87334	chaperonin, 60 kDa
32	1791	60.3	541	2	S61301	heat shock protein
33	1789.5	60.3	541	2	JN0511	heat shock protein
34	1788.5	60.3	544	2	S61303	heat shock protein
35	1786	60.2	545	2	C70489	GroEL - Aquifex ae
36	1780.5	60.0	544	2	JC6063	chaperonin groEL -
37	1780.5	60.0	544	2	B83720	class I heat-shock
38	1774.5	59.8	545	2	S51563	heat shock protein
39	1773	59.7	544	1	B43827	chaperonin groEL -
40	1765	59.5	546	2	S34938	heat shock protein
41	1761.5	59.3	544	2	B82048	chaperonin, 60 Kd
42	1761	59.3	548	2	AG0043	60 kDa chaperonin
43	1759	59.3	548	2	C64076	chaperonin groEL -
44	1755.5	59.1	539	2	B49855	heat shock protein
45	1755	59.1	545	2	G81328	60 kD chaperonin (

ALIGNMENTS

RESULT 1
JC2562
chaperonin groELx protein - Amoeba proteus
C;Species: Amoeba proteus
C;Date: 17-May-1995 #sequence_revision 14-Jul-1995 #text_change 26-Aug-1999
C;Accession: JC2562
R;Ahn, T.I.; Lim, S.T.; Leeu, H.K.; Lee, J.E.; Jeon, K.W.
Gene 148[128], 43-49, 1994
A;Title: A novel strong promoter of the groEx operon of symbiotic bacteria in Amoeba prot
A;Reference number: JC2561
A;Note: due to a typographical error the volume number 148 appears as 128
A;Accession: JC2562
A;Molecule type: DNA
A;Residues: 1-551 <AHN>
A;Cross-references: GB:M86549; NID:g155400; PIDN:AAC09381.1; PID:g155402
C;Comment: This protein is involved in the assembly of oligomeric protein complexes, and
C;Genetics:
A;Gene: groELx
C;Superfamily: chaperonin groEL
C;Keywords: molecular chaperone

Alignment Scores:
Pred. No.: 8.77e-95 Length: 551
Score: 1866.00 Matches: 373
Percent Similarity: 81.42% Conservative: 74
Best Local Similarity: 67.94% Mismatches: 96
Query Match: 62.87% Indels: 6
DB: 2 Gaps: 2

US-09-077-574A-1 (1-1647) x JC2562 (1-551)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||:::   |||         ||||||:::::::::     ||||||:::
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgGlnGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::   ||||||   |||:::|||||   |||||||||||||||:::|||:::
Db         22 LeuAlaAspArgValLysAlaThrMetGlyProSerGlyArgAsnValValLeuGluArg 41
```

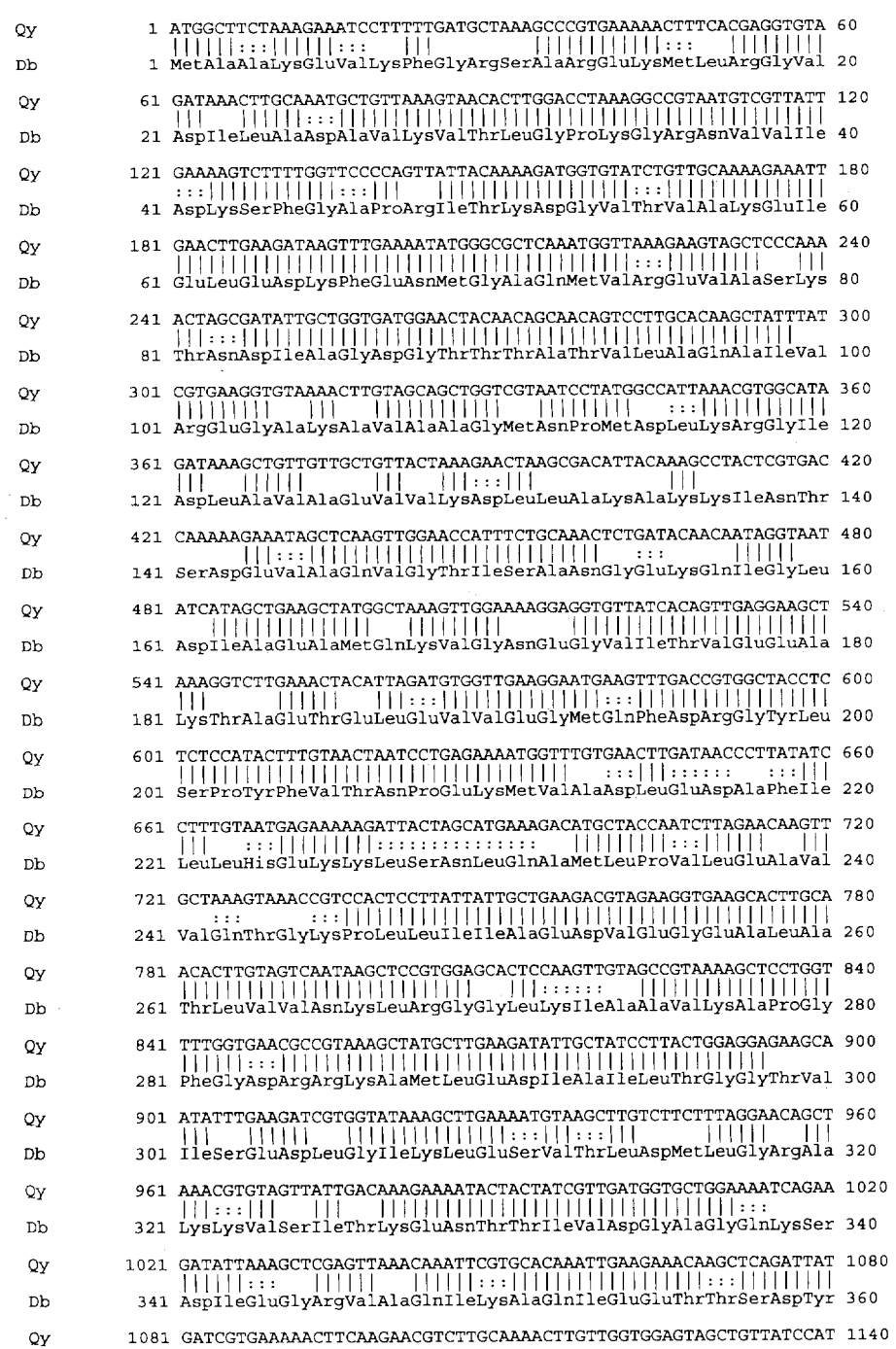
```
Qy       127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db        42 SerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy       187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db        62 GluAsnArgPheLysAsnMetGlyAlaGlnMetValLysGluValAlaAlaLysThrSer 81

Qy       247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db        82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleValValGlu 101

Qy       307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db       102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy       367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db       122 AlaValThrAlaIleThrLysGluLeuGlnLysMetSerLysProCysLysAspGlyLys 141

Qy       427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db       142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGlnAlaIleGlySerIleIle 161

Qy       487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db       162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy       547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db       182 LeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy       607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db       202 TyrPheIleAsnAsnGlnGlnAsnMetSerAlaGluLeuGluHisProPheIleLeuLeu 221

Qy       667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db       222 ValAspLysLysIleAlaThrIleArgAspMetLeuSerValLeuGluAlaValAlaLys 241

Qy       727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db       242 SerGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261

Qy       787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db       262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy       847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db       282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrAsnGlyGlnValIleSer 301

Qy       907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db       302 GluGluIleGlyThrSerLeuGluThrAlaSerLeuGluSerLeuGlyThrAlaLysArg 321

Qy       967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db       322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy      1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db       342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrSerSerAspTyrAspArg 361

Qy      1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db       362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy      1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db       382 AlaAlaThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy      1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db       402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy      1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db       422 GlnLysValLeuAspGlyLeuLysGlyAspAsnAlaAspGlnAspMetGlyIleAsnIle 441

Qy      1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db       442 LeuArgArgAlaIleGluSerProLeuArgGlnIleValAlaAsnAlaGlyTyrGluSer 461

Qy      1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db       462 SerValIleValAsnLysValAlaGluHisLysAspAsnPheGlyPheAsnAlaAlaThr 481

Qy      1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db       482 GlyGlnTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgThr 501

Qy      1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db       502 AlaLeuGlnAsnAlaAlaSerValArgSerLeuMetLeuThrThrGluCysMetValAla 521

Qy      1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC----------GGTGGTATG 1614
Db       522 AspLeu------ProLysLysAspGluGlyMetAlaGlyAlaGlyAspMetGlyGlyMet 539

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db       540 GlyGlyMetGlyGlyMetGlyGlyMet 548
```

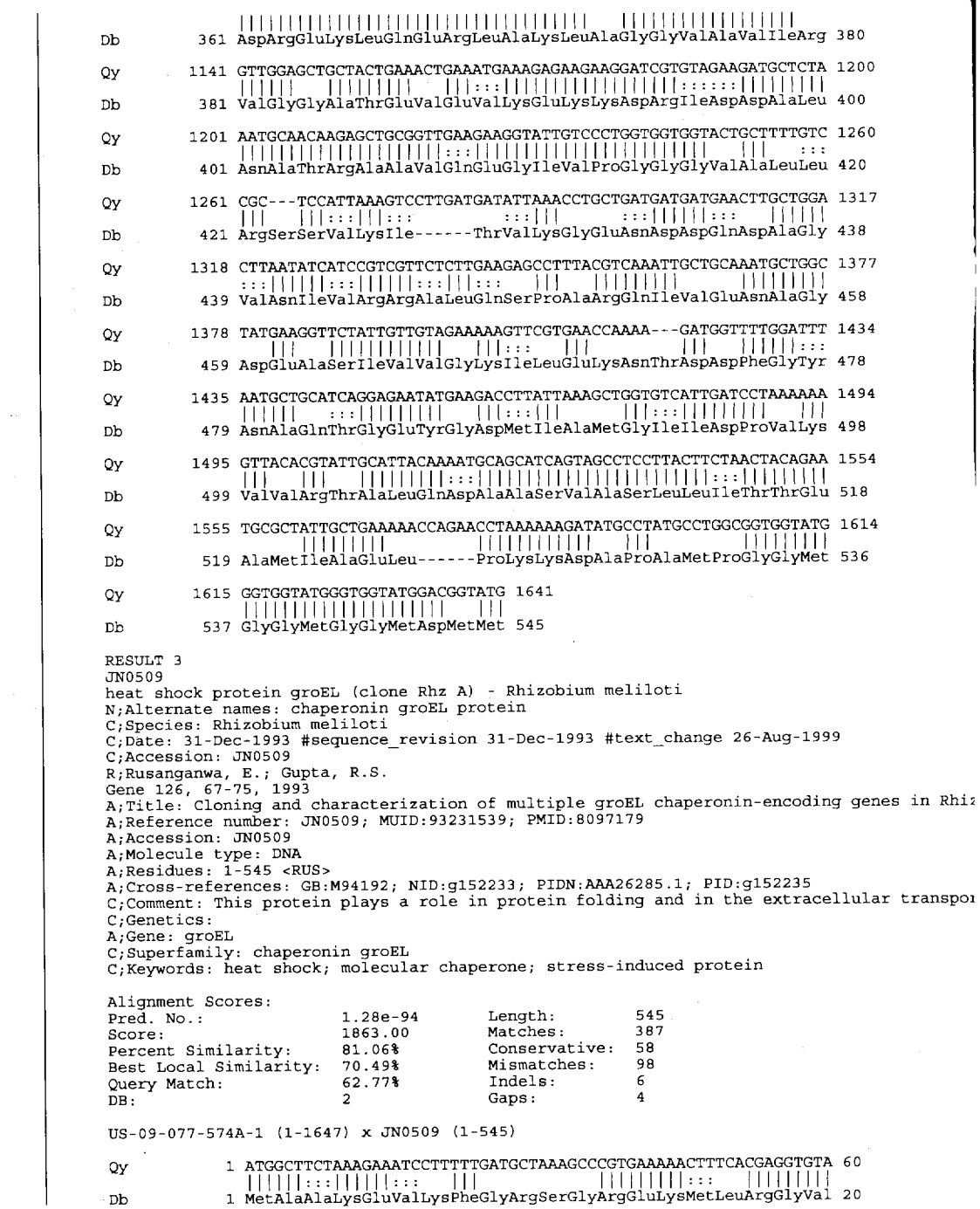
RESULT 2
C95311
groEL2 chaperonin [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: C95311
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: C95311
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-545 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK65053.1; PID:g14523485; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: groEL2
A;Genome: plasmid
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1.13e-94	Length:	545
Score:	1864.00	Matches:	387
Percent Similarity:	81.24%	Conservative:	59
Best Local Similarity:	70.49%	Mismatches:	97
Query Match:	62.80%	Indels:	6
DB:	2	Gaps:	4

US-09-077-574A-1 (1-1647) x C95311 (1-545)





RESULT 3
JN0509
heat shock protein groEL (clone Rhz A) - Rhizobium meliloti
N;Alternate names: chaperonin groEL protein
C;Species: Rhizobium meliloti
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C;Accession: JN0509
R;Rusanganwa, E.; Gupta, R.S.
Gene 126, 67-75, 1993
A;Title: Cloning and characterization of multiple groEL chaperonin-encoding genes in Rhiz
A;Reference number: JN0509; MUID:93231539; PMID:8097179
A;Accession: JN0509
A;Molecule type: DNA
A;Residues: 1-545 <RUS>
A;Cross-references: GB:M94192; NID:g152233; PIDN:AAA26285.1; PID:g152235
C;Comment: This protein plays a role in protein folding and in the extracellular transpor
C;Genetics:
A;Gene: groEL
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone; stress-induced protein

Alignment Scores:

Pred. No.:	1.28e-94	Length:	545
Score:	1863.00	Matches:	387
Percent Similarity:	81.06%	Conservative:	58
Best Local Similarity:	70.49%	Mismatches:	98
Query Match:	62.77%	Indels:	6
DB:	2	Gaps:	4

US-09-077-574A-1 (1-1647) x JN0509 (1-545)

```
Qy        61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy       121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy       181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       121 AspLeuAlaValAlaGluValValLysAspLeuLeuAlaLysAlaLysLysIleAsnThr 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       141 SerAspGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluLysGlnIleGlyLeu 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       161 AspIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaPheIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       241 ValGlnThrGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy      1261 CGC---TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db       421 ArgSerSerValLysIle------ThrValLysGlyGluAsnAspAspGlnAspAlaGly 438

Qy      1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db       439 ValAsnIleValArgArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458

Qy      1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAA---GATGGTTTTGGATTT 1434
Db       459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspAspPheGlyTyr 478

Qy      1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db       479 AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498

Qy      1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db       499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518

Qy      1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATG 1614
Db       519 AlaMetIleAlaGluLeu------ProLysLysAspAlaProAlaMetProGlyGlyMet 536

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db       537 GlyGlyMetGlyGlyMetAspMetMet 545
```

RESULT 4
F82783
60kDa chaperonin XF0615 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: F82783
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <SIM>
A;Cross-references: GB:AE003907; GB:AE003849; NID:g9105484; PIDN:AAF83425.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0615
C;Superfamily: chaperonin groEL

Alignment Scores:
Pred. No.: 1.36e-94 Length: 547

Score:	1862.50	Matches:	366
Percent Similarity:	81.39%	Conservative:	80
Best Local Similarity:	66.79%	Mismatches:	97
Query Match:	62.75%	Indels:	5
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x F82783 (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluIleIlePheSerGluLysAlaArgSerArgMetValHisGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlySerProIleIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspHisAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValArgValSerLysGluAsnThrThrIleIleAspGlyIleGlyAspAsnAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCT---------GGCGGTGGT 1611
Db        521 ValAlaGlu------AlaProLysLysAspGluProThrProProAlaAlaGlyGlyGly 538

Qy       1612 ATGGGTGGTATGGGTGGTATGGAC 1635
Db        539 MetGlyGlyMetGlyGlyMetAsp 546
```

RESULT 5
B47073
chaperonin GroEL - Chromatium vinosum
C;Species: Chromatium vinosum
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C;Accession: B47073
R;Ferreyra, R.G.; Soncini, F.C.; Viale, A.M.
J. Bacteriol. 175, 1514-1523, 1993
A;Title: Cloning, characterization, and functional expression in Escherichia coli of cha
A;Reference number: A47073; MUID:93186721; PMID:8444812
A;Accession: B47073
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-546 <FER>
A;Cross-references: GB:M99443; NID:g145007; PIDN:AAA23319.1; PID:g145009
A;Note: sequence extracted from NCBI backbone (*NCBIN:126968, NCBIP:126970*)
C;Superfamily: chaperonin groEL
C;Keywords: molecular chaperone

Alignment Scores:

Pred. No.:	4.79e-94	Length:	546
Score:	1852.50	Matches:	374
Percent Similarity:	81.02%	Conservative:	70
Best Local Similarity:	68.25%	Mismatches:	101
Query Match:	62.42%	Indels:	3
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x B47073 (1-546)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetSerAlaLysAspValLysPheGlyGlyAspAlaArgValArgMetMetGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluLysSerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrSerAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaMetVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyMet 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValGluAlaAlaThrGluGluLeuLysLysLeuSerLysProCysProArg 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 ProMetAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspAspSerIleGlyThr 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 ThrSerLeuGlnAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnSerAlaGluLeuAspAlaProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuTyrAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlyVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnThrIleArgGlyIleValLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyAlaThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuThrAspLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysArgValGlnValGlyLysAspGluThrThrIleIleAspGlySerGlySerGluIle 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AspIleLysAlaArgCysGluGlnIleArgAlaGlnValGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaIleAlaAlaValLysAspLeuLysGlyAlaAsnHisAspGlnAspValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 AlaIleAlaArgArgAlaMetGluGluProLeuArgGlnIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValIleLeuHisLysValAlaGluGlyThrGlyAsnPheGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaAsnGlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnAsnSerCysSerValAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATGGGT 1617
Db        521 IleAlaAsp------GluProLysAspAspAlaProAlaMetProGlyGlyGlyMetGly 538

Qy       1618 GGTATGGGTGGTATGGACGGTATG 1641
Db        539 AspMetGlyGlyMetGlyMetMet 546
```

RESULT 6
A41468
60K heat shock protein htpB - Legionella pneumophila
C;Species: Legionella pneumophila
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C;Accession: A41468; A41472
R;Hoffman, P.S.; Houston, L.; Butler, C.A.
Infect. Immun. 58, 3380-3387, 1990
A;Title: Legionella pneumophila htpAB heat shock operon: nucleotide sequence and express
A;Reference number: A41468; MUID:90382960; PMID:2205580
A;Accession: A41468
A;Molecule type: DNA
A;Residues: 1-550 <HOF>
A;Cross-references: GB:M31918
R;Sampson, J.S.; O'Connor, S.P.; Holloway, B.P.; Plikaytis, B.B.; Carlone, G.M.; Mayer,
Infect. Immun. 58, 3154-3157, 1990
A;Title: Nucleotide sequence of htpB, the Legionella pneumophila gene encoding the 58-ki
A;Reference number: A41472; MUID:90354095; PMID:2117582
A;Accession: A41472
A;Molecule type: DNA
A;Residues: 3-188,'S',190-195,'LIAVH',202-347,'T',349-550 <SAM>
A;Cross-references: GB:M91673; GB:M35149; NID:g149689; PIDN:AAA25298.1; PID:g149690
C;Genetics:
A;Gene: htpB
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1.23e-93	Length:	550
Score:	1845.00	Matches:	367
Percent Similarity:	82.02%	Conservative:	80
Best Local Similarity:	67.34%	Mismatches:	96

Query Match: 62.16% Indels: 2
DB: 2 Gaps: 1

US-09-077-574A-1 (1-1647) x A41468 (1-550)

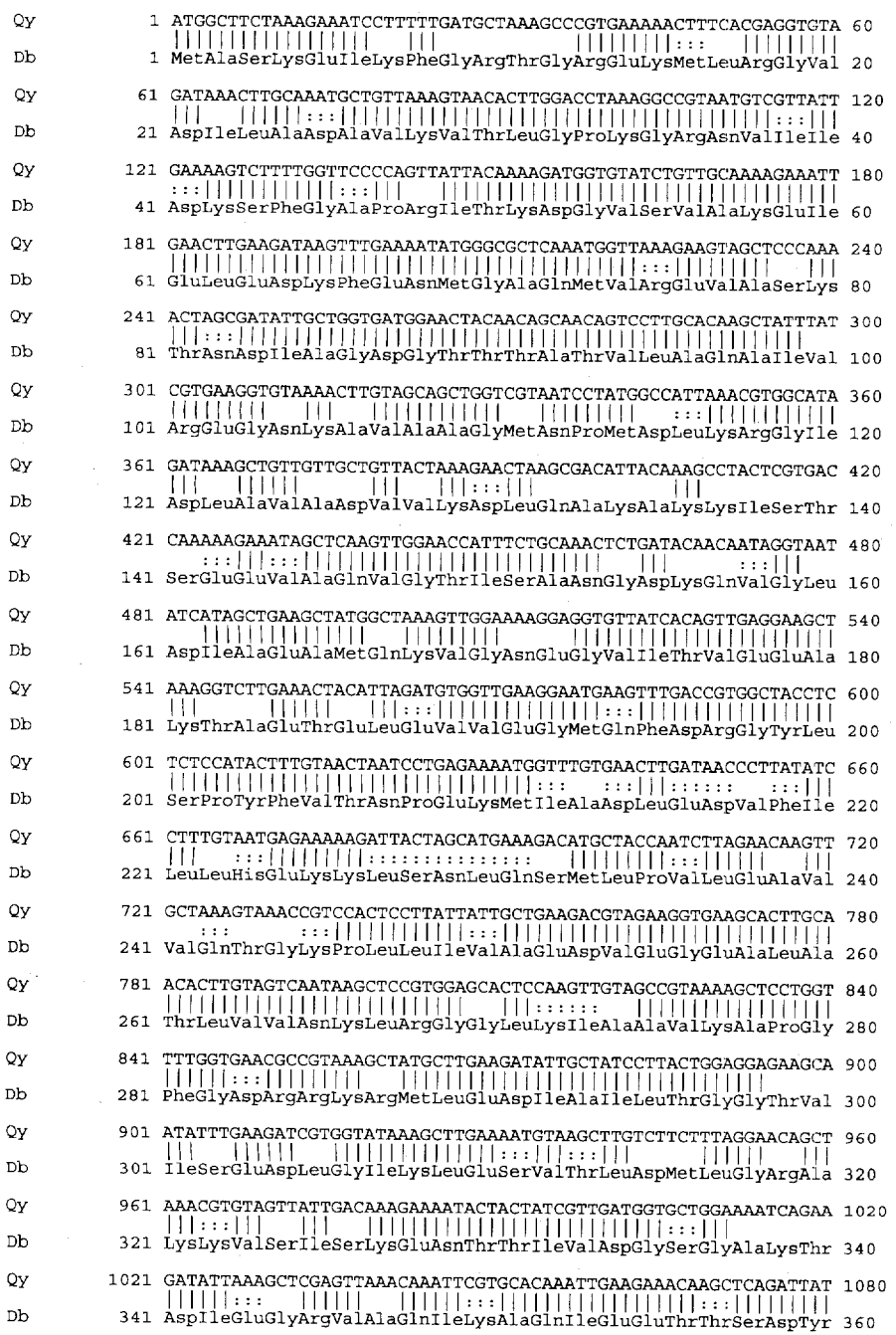
```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||:::   |||         ||||||   ::::::      ||||||:::
Db          4 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 23

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::||||||:::|||||||:::||||||:::|||||||||||||||||:::||||||
Db         24 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 43

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   :::||||||||||||||||||||||||||||||||||||||
Db         44 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 63

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              |||   :::|||   |||||||||||||||||||||||||||||||   |||||||||
Db         64 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 83

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||:::::|||      |||
Db         84 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 103

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||   |||   ||||||||||   |||||||||   :::|||||||||||||||||||
Db        104 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 123

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||:::|||||||||||||:::|||       :::::::||||||   :::|||   |||
Db        124 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 143

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 163

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              ||||||||||||   ||||||||||||   ||||||||||||||||:::            |||
Db        164 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 183

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              ||||||      |||   ||||||||||||||:::||||||||||||||||:::||||||
Db        184 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 203

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||:::   |||   :::   |||   ||||||||:::::|||:::||||||
Db        204 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 223

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::|||||:::::|||:::::::::||||||   :::||||||   |||||||||
Db        224 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 243

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 SerGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 263

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              |||||||||   :::||||||   ::::::|||   |||||||||||||||||||||||
Db        264 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 283

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::||||||||||||||||:::||||||||||||||||||   |||:::   |||
Db        284 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 303

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||       ||||||       :::|||       ||||||:::|||||||||
Db        304 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 323

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||:::   ||||||||||||||||||:::||||||   ||||||:::   :::|||
Db        324 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 343

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              ||||||:::   ||||||||||||||||:::|||||||||:::||||||||||||||||
Db        344 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 363

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||:::|||||||||   ||||||||||||||||||   ||||||
Db        364 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 383

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||||   ||||||||||||||||||   ||||||||||||||||||:::|||
Db        384 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 403

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        404 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 423

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              |||   ||||||   :::|||       :::||||||:::       |||:::||||||
Db        424 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 443

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::||||||:::::::|||   |||:::||||||||||       ||||||||||||||
Db        444 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrArgAsnAlaGlyTyrGluAla 463

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::||||||   ||||||   |||   ||||||   :::|||||||||||||:::
Db        464 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 483

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||||||||   |||:::::::::   |||:::::||||||   ||||||||||||:::
Db        484 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 503

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||||||||||||||||||||:::|||||||||||||||   :::|||
Db        504 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 523

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
              :::         |||||||||:::      :::      |||   ||||||||||||||
Db        524 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 541

Qy       1627 GGTATGGACGGTATG 1641
              ||||||   ||||||
Db        542 GlyMetGlyGlyMet 546
```

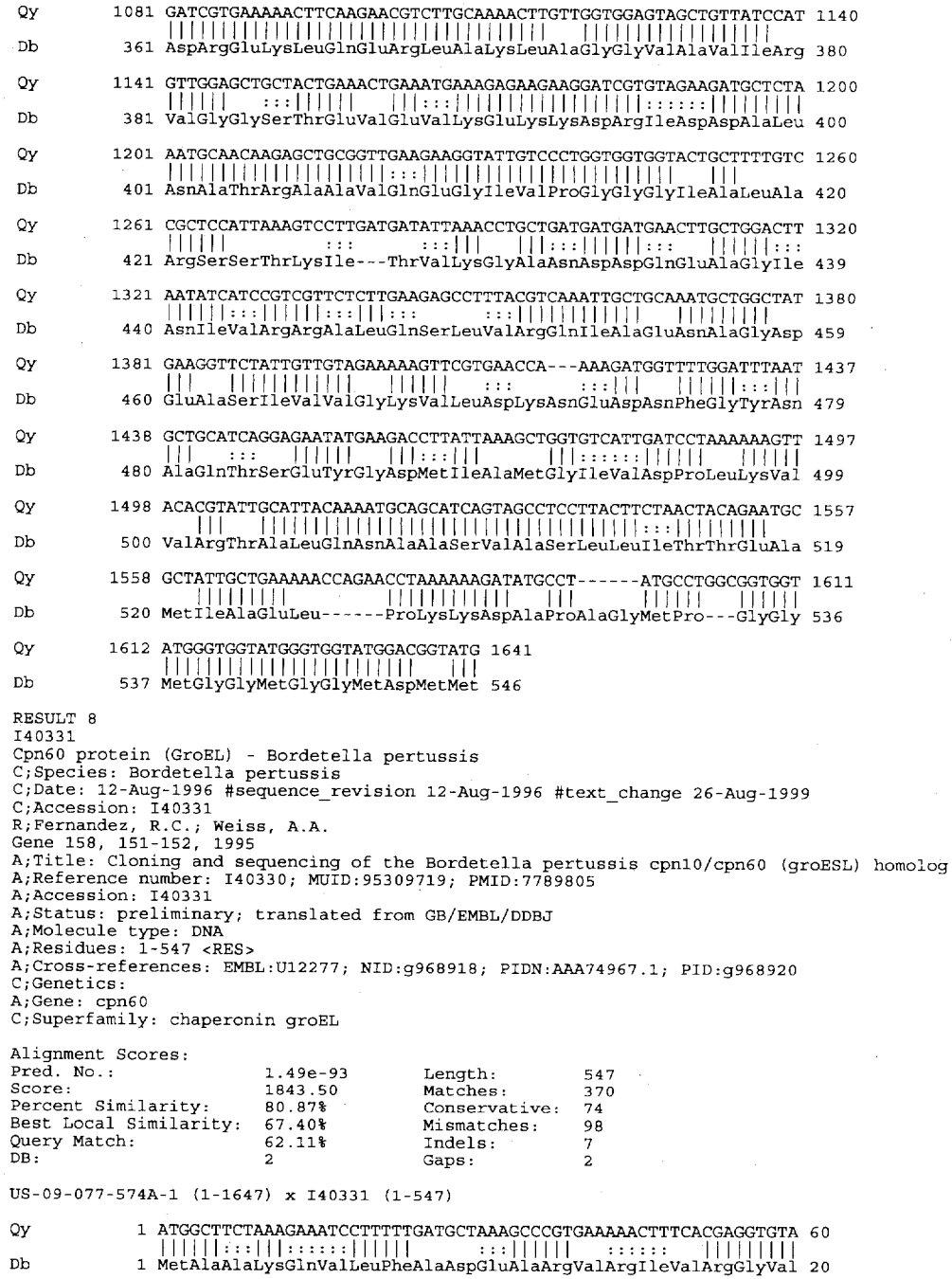
RESULT 7
S65596
heat shock protein 60 - Rhizobium leguminosarum
N;Alternate names: chaperonin 60
C;Species: Rhizobium leguminosarum
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C;Accession: S65596
R;Wallington, E.J.; Lund, P.A.
Microbiology 140, 113-122, 1994
A;Title: Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes.
A;Reference number: S65596; MUID:94214663; PMID:7909257
A;Accession: S65596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <WAL>
A;Cross-references: EMBL:L20775; NID:g387873; PIDN:AAA26246.1; PID:g387874
C;Genetics:
A;Gene: cpn60
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone

Alignment Scores:

Pred. No.:	1.31e-93	Length:	546
Score:	1844.50	Matches:	384
Percent Similarity:	80.91%	Conservative:	61
Best Local Similarity:	69.82%	Mismatches:	98
Query Match:	62.15%	Indels:	7
DB:	2	Gaps:	5

US-09-077-574A-1 (1-1647) x S65596 (1-546)



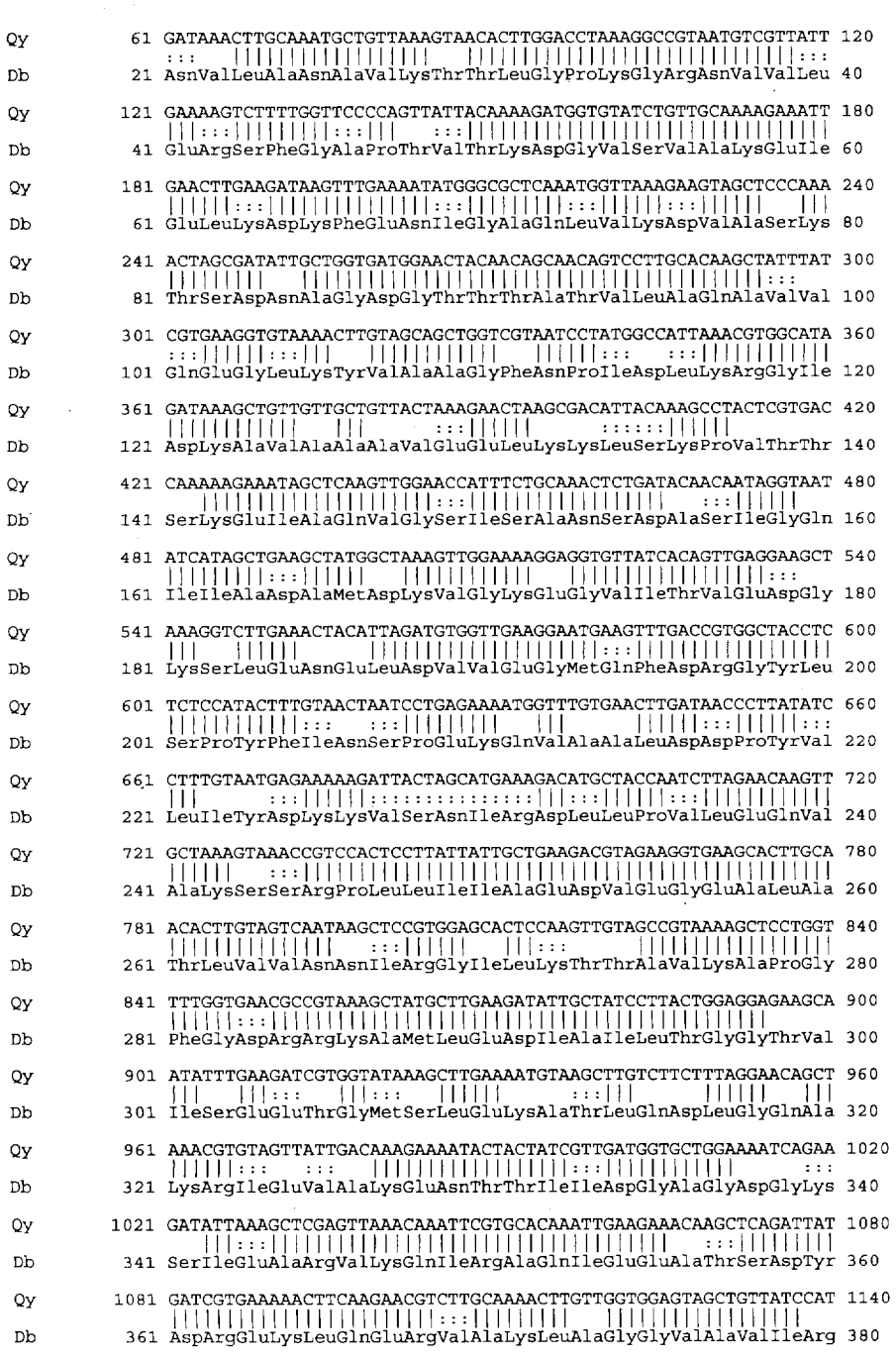


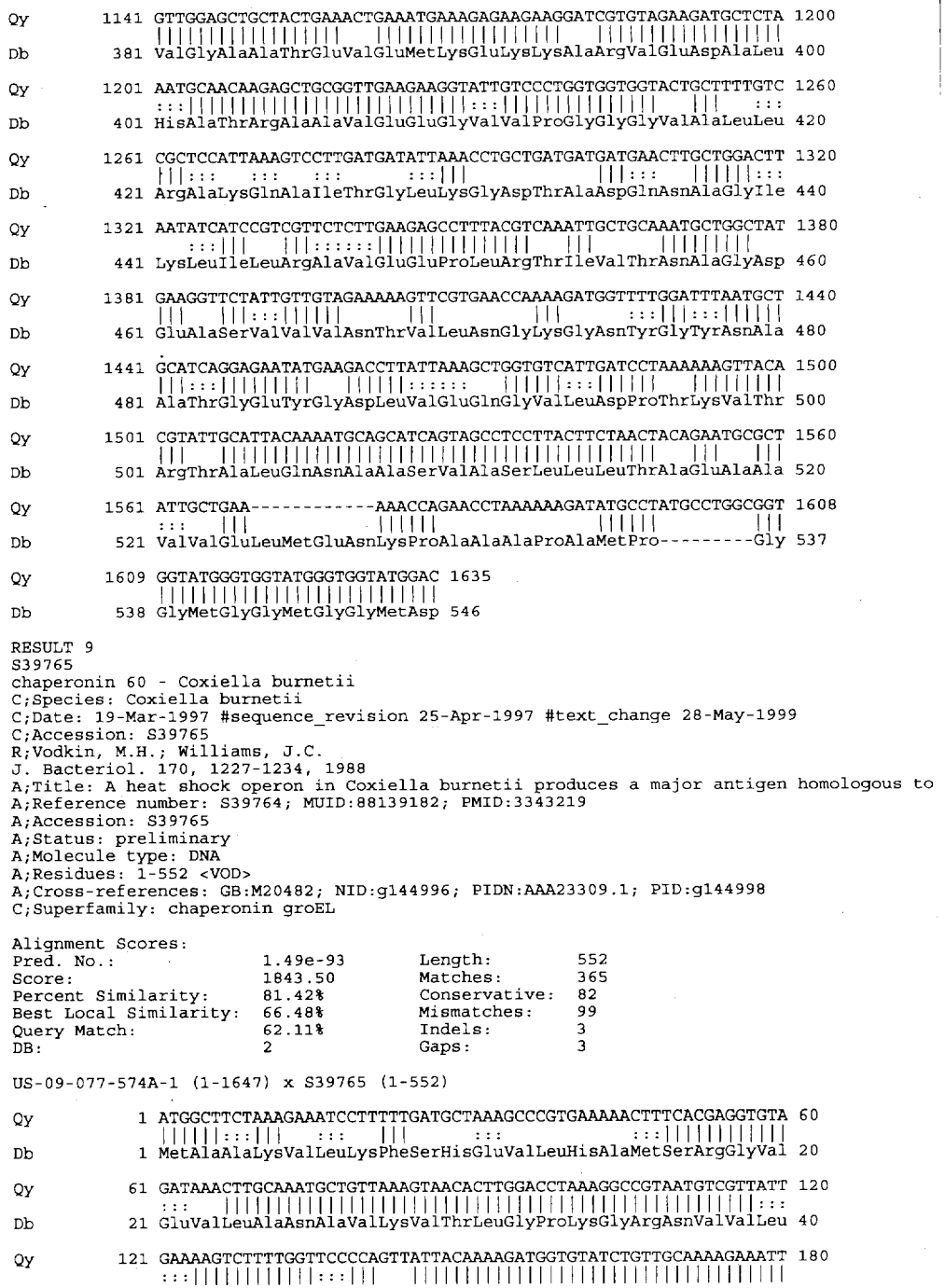
RESULT 8
I40331
Cpn60 protein (GroEL) - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 26-Aug-1999
C;Accession: I40331
R;Fernandez, R.C.; Weiss, A.A.
Gene 158, 151-152, 1995
A;Title: Cloning and sequencing of the Bordetella pertussis cpn10/cpn60 (groESL) homolog
A;Reference number: I40330; MUID:95309719; PMID:7789805
A;Accession: I40331
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-547 <RES>
A;Cross-references: EMBL:U12277; NID:g968918; PIDN:AAA74967.1; PID:g968920
C;Genetics:
A;Gene: cpn60
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1.49e-93	Length:	547
Score:	1843.50	Matches:	370
Percent Similarity:	80.87%	Conservative:	74
Best Local Similarity:	67.40%	Mismatches:	98
Query Match:	62.11%	Indels:	7
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x I40331 (1-547)





RESULT 9
S39765
chaperonin 60 - Coxiella burnetii
C;Species: Coxiella burnetii
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 28-May-1999
C;Accession: S39765
R;Vodkin, M.H.; Williams, J.C.
J. Bacteriol. 170, 1227-1234, 1988
A;Title: A heat shock operon in Coxiella burnetii produces a major antigen homologous to
A;Reference number: S39764; MUID:88139182; PMID:3343219
A;Accession: S39765
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-552 <VOD>
A;Cross-references: GB:M20482; NID:g144996; PIDN:AAA23309.1; PID:g144998
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1.49e-93	Length:	552
Score:	1843.50	Matches:	365
Percent Similarity:	81.42%	Conservative:	82
Best Local Similarity:	66.48%	Mismatches:	99
Query Match:	62.11%	Indels:	3
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x S39765 (1-552)

```
Db        41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy       181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :::
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
             ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 ThrSerAspAspAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleLeu 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
               ||||||:::|||   |||  ||||||  |||||||||   :::|||||||||||||
Db       101 ValGluGlyIleLysAlaValIleAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
             ||||||||||||   |||          ||||||      |||:::||||||   :::|||
Db       121 AspLysAlaValThrAlaAlaValAlaGluLeuLysLysIleSerLysProCysLysAsp 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
             ||||||  ||||||||||||||||||||||||||||||||  :::||||||:::
Db       141 GlnLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspLysSerIleGlyAsp 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
             ||||||||||||||||  ||||||||||||  |||||||||||||||||||:::
Db       161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
             ||||||||      |||:::||||||||||||||:::|||||||||||||||||||
Db       181 SerGlyLeuGluAsnAlaLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
             ||||||||||||:::  |||  :::  |||      ||||||:::||||||:::|||
Db       201 SerProTyrPheIleAsnAsnGlnGlnAsnMetSerAlaGluLeuGluAsnProPheIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
             |||      :::|||||||||:::::::::::::::::::::|||:::||||||   |||
Db       221 LeuLeuValAspLysLysIleSerAsnIleArgGluLeuIleProLeuLeuGluAsnVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
             ||||||      ||||||||||||:::||||||||||||:::||||||||||||||||||
Db       241 AlaLysSerGlyArgProLeuLeuValIleAlaGluAspIleGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
             ||||||||||||||   :::||||||  ::::::|||  ||||||||||||||||||
Db       261 ThrLeuValValAsnAsnIleArgGlyValValLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
             ||||||:::|||||||||||||||||:::|||||||||:::|||||||||||||:::
Db       281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaValLeuThrGlyGlyLysVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
             |||   |||:::   |||:::  ||||||      ||||||      ||||||:::|||
Db       301 IleSerGluGluValGlyLeuSerLeuGluAlaAlaSerLeuAspAspLeuGlySerAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
             ||||||||||:::   |||::::::||||||||:::||||||:::|||   :::
Db       321 LysArgValValValThrLysAspAspThrThrIleIleAspGlySerGlyAspAlaGly 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
             |||||||||   ||||||:::|||||||||   :::|||||   :::|||||||||||||
Db       341 AspIleLysAsnArgValGluGlnIleArgLysGluIleGluAsnSerSerSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
             |||:::|||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
             ||||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||
Db       381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
             :::|||||||||||||||||||||||||:::|||||||||||||||||   |||   :::
Db       401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuIle 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
             |||   :::|||   ||||||   ::::::      ::::::|||:::      |||:::
Db       421 ArgValLeuLysSerLeuAspSerValGluValGluAsnGluAspGlnArgValGlyVal 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
               |||   ||||||::::::     ||||||   ||||||      |||   |||
Db       441 GluIleAlaArgArgAlaMetAlaTyrProLeuSerGlnIleValLysAsnThrGlyVal 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
               :::   :::::::|||   :::||||||         ||||||      :::|||:::|||
Db       461 GlnAlaAlaValValAlaAspLysValLeuAsnHisLysAspValAsnTyrGlyTyrAsn 480

Qy      1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
             ||||||:::|||||||||   |||:::|||:::   |||::::::||||||   ||||||
Db       481 AlaAlaThrGlyGluTyrGlyAspMetIleGluMetGlyIleLeuAspProThrLysVal 500

Qy      1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
             ||||||   ||||||||||||||||||||:::|||   |||:::::|||||||||||||
Db       501 ThrArgThrAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluCys 520

Qy      1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGT---ATG 1614
               :::   |||   |||:::   ||||||:::      ||||||||||||||   |||
Db       521 MetValThrGluAlaProLys---LysLysGluGluSerMetProGlyGlyGlyAspMet 539

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
             ||||||||||||||||||   ||||||
Db       540 GlyGlyMetGlyGlyMetGlyGlyMet 548
```

RESULT 10
F95967
probable heat shock protein groEL [imported] - Sinorhizobium meliloti (strain 1021) maga
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: F95967
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: F95967
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49406.1; PID:g15140892; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: groEL5; SMb21566
A;Genome: plasmid
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	2.04e-93	Length:	542
Score:	1841.00	Matches:	373
Percent Similarity:	81.32%	Conservative:	71
Best Local Similarity:	68.32%	Mismatches:	96
Query Match:	62.03%	Indels:	6
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x F95967 (1-542)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::||||||:::   |||         |||||||||:::::   |||||||||
Db          1 MetAlaAlaLysGluValLysPheGlnThrAspAlaArgGluArgMetLeuArgGlyVal 20
```

```
Qy        61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        21 AspValLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy       121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy       181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerArg 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        81 ThrAsnAspLeuAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       101 ArgGluGlyAlaLysAlaValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       121 AspLeuAlaValAspAlaValValLysGluLeuLysAsnAsnAlaArgLysIleSerLys 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       141 AsnSerGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAspThrGluIleGlyArg 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       161 TyrLeuAlaGluAlaMetGluLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       201 SerProTyrPheIleThrAsnGlnAspLysMetArgValGluLeuGluAspProTyrIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       221 LeuIleHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       241 ValGlnSerGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       301 ValSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       321 LysLysValSerIleGluLysGluAsnThrThrIleIleAspGlyAlaGlySerLysAla 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       341 AspIleGluGlyArgThrAlaGlnIleArgAlaGlnIleGluGluThrThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       401 HisAlaThrArgAlaAlaValGluGluGlyIleLeuProGlyGlyGlyValAlaLeuLeu 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       421 ArgAlaValLysAlaLeuAspGlyLeuLysThrAlaAsnAsnAspGlnArgValGlyVal 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       441 AspLeuValArgArgAlaIleGluAlaProValArgGlnIleAlaGluAsnAlaGlyAla 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
Db       461 GluGlySerIleIleValGlyLysLeuArgGluLysThrGluPheSerTyrGlyTrpAsn 480

Qy      1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       481 AlaGlnThrAsnGluTyrGlyAspLeuTyrAlaMetGlyValIleAspProAlaLysVal 500

Qy      1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 520

Qy      1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db       521 MetIleAlaGluLysProLysLysGluAlaAlaProAlaLeuProAlaGly--------- 537

Qy      1618 GGTATGGGTGGTATGGAC 1635
Db       538 ------GlyGlyMetAsp 541
```

RESULT 11
AD2660
60 KDA chaperonin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AD2660
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClella
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, I
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AD2660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41698.1; PID:g17739044; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: groEL
A;Map position: circular chromosome
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	2.17e-93	Length:	544
Score:	1840.50	Matches:	376
Percent Similarity:	81.95%	Conservative:	69
Best Local Similarity:	69.24%	Mismatches:	95
Query Match:	62.01%	Indels:	3
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x AD2660 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuLysGlyVal 20
```

```
Qy        61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy       121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy       181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       121 AspLeuAlaValAlaGluValValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       241 ValGlnThrGlyLysProLeuValIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspAspGlnGluAlaGlyIle 439

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCA---AAAGATGGTTTTGGATTTAAT 1437
Db       460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479

Qy      1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy      1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 519

Qy      1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
Db       520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet 539

Qy      1615 GGTGGTATG 1623
Db       540 GlyGlyMet 542
```

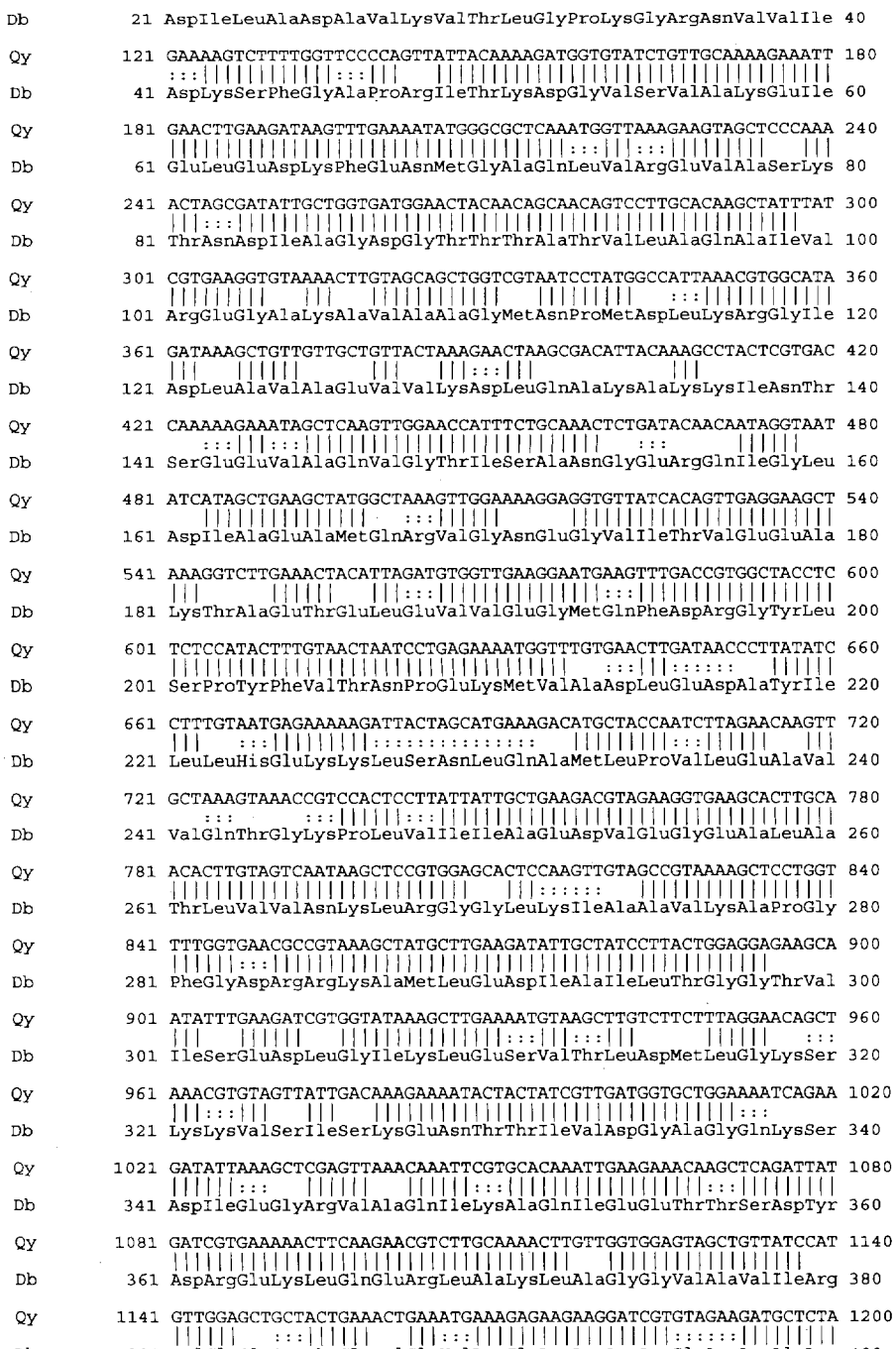
RESULT 12
B97442
60K chaperonin (protein cpn60) (groEL protein) [imported] - Agrobacterium tumefaciens (s
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: B97442
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B97442
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86491.1; PID:g15155641; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1220
A;Map position: circular chromosome
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	2.17e-93	Length:	544
Score:	1840.50	Matches:	376
Percent Similarity:	81.95%	Conservative:	69
Best Local Similarity:	69.24%	Mismatches:	95
Query Match:	62.01%	Indels:	3
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x B97442 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuLysGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
```

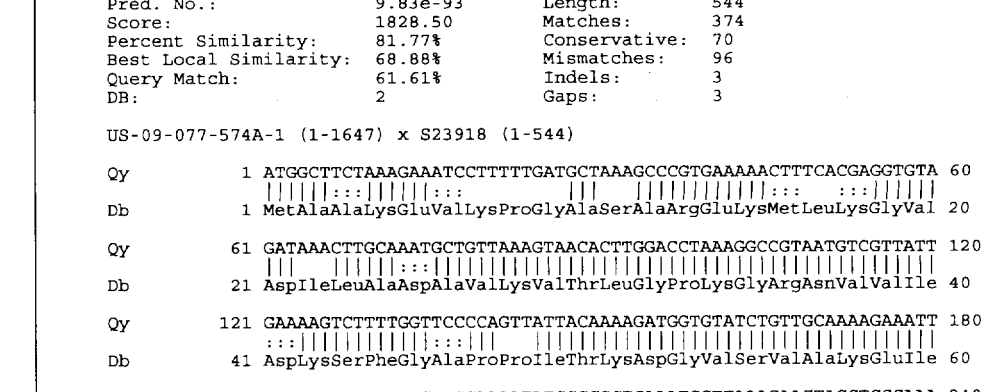


RESULT 13
S23918
groEL protein - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C;Accession: S23918
R;Segal, G.; Ron, E.Z.
submitted to the EMBL Data Library, August 1992
A;Description: Cloning and sequencing of the GroE operon of Agrobacterium tumefaciens C58
A;Reference number: S23917
A;Accession: S23918
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <SEG>
A;Cross-references: EMBL:X68263; NID:g1019913; PIDN:CAA48331.1; PID:g39095
C;Superfamily: chaperonin groEL

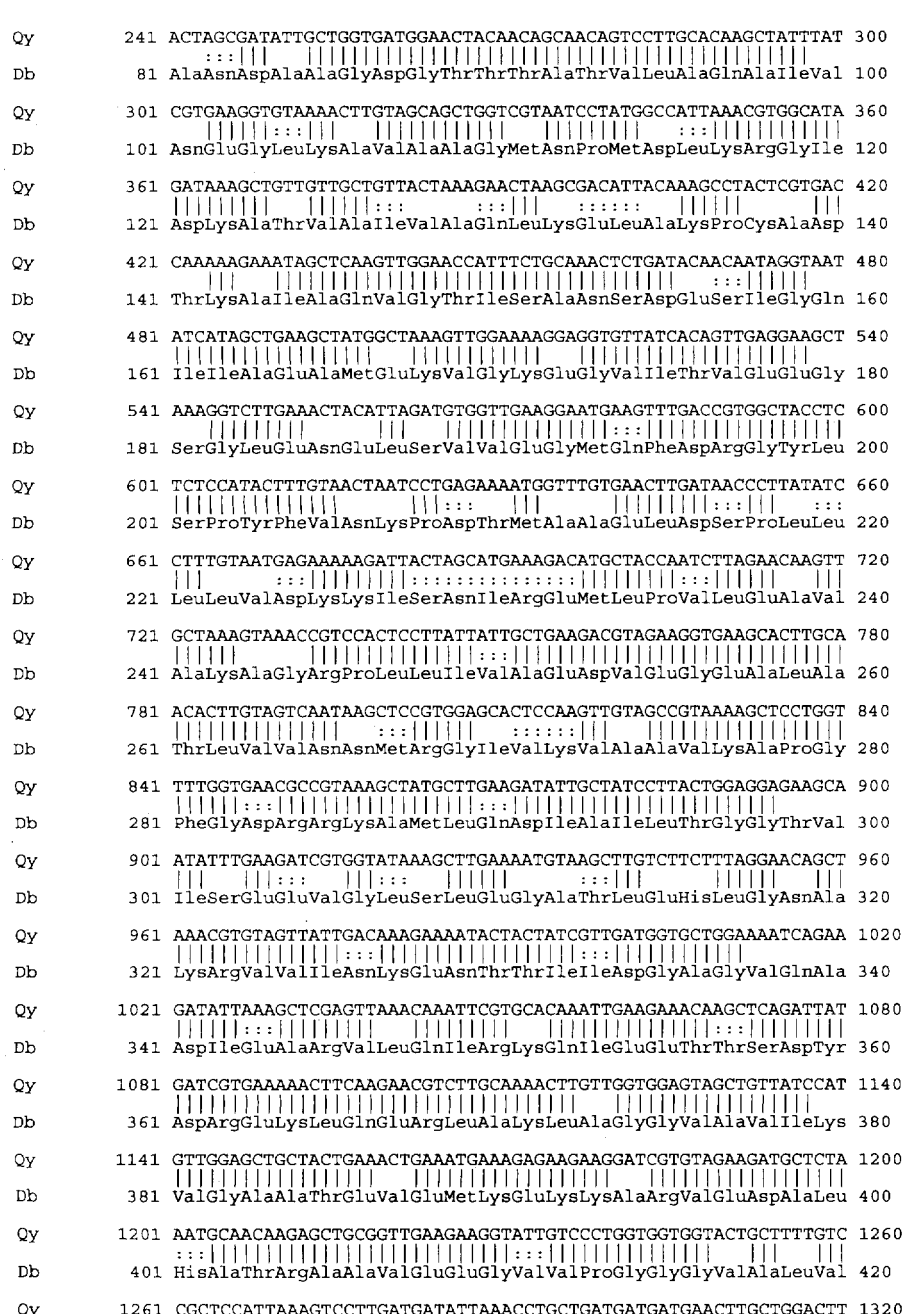
Alignment Scores:

Pred. No.:	9.83e-93	Length:	544
Score:	1828.50	Matches:	374
Percent Similarity:	81.77%	Conservative:	70
Best Local Similarity:	68.88%	Mismatches:	96
Query Match:	61.61%	Indels:	3
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x S23918 (1-544)



```
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       121 AspLeuAlaValAlaGluValValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       241 ValGlnThrGlyLysProLeuValIleIleAlaGluAspValGluGlyGluArgLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db       421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspAspGlnGluAlaGlyIle 439

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCA---AAAGATGGTTTTGGATTTAAT 1437
Db       460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479

Qy      1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db       480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy      1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db       500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerValLeuIleThrThrGluAla 519

Qy      1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
Db       520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet 539

Qy      1615 GGTGGTATG 1623
Db       540 GlyGlyMet 542

RESULT 14
B83098
GroEL protein PA4385 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83098
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83098
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <STO>
A;Cross-references: GB:AE004854; GB:AE004091; NID:g9950606; PIDN:AAG07773.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: groEL; PA4385
C;Superfamily: chaperonin groEL

Alignment Scores:
Pred. No.:              1.43e-92        Length:         547
Score:                  1825.50         Matches:        367
Percent Similarity:     80.51%          Conservative:   75
Best Local Similarity:  66.85%          Mismatches:     102
Query Match:            61.51%          Indels:         5
DB:                     2               Gaps:           2

US-09-077-574A-1 (1-1647) x B83098 (1-547)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80
```

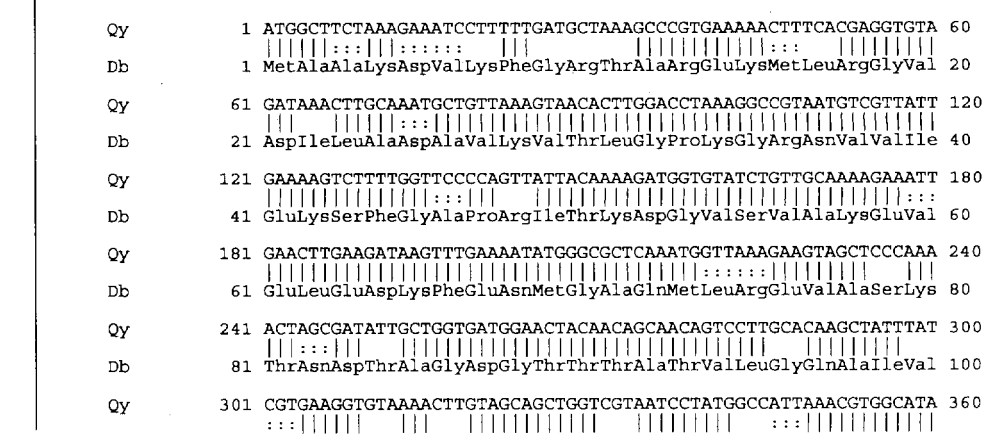


RESULT 15
S22347
groEL - Brucella abortus
C;Species: Brucella abortus
C;Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
C;Accession: S22347
R;Gor, D.; Mayfield, J.E.
Biochim. Biophys. Acta 1130, 120-122, 1992
A;Title: Cloning and nucleotide sequence of the Brucella abortus groE operon.
A;Reference number: S22346; MUID:92182006; PMID:1347461
A;Accession: S22347
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <GOR>
A;Cross-references: EMBL:M82975; NID:g144109; PIDN:AAA22997.1; PID:g144111
C;Superfamily: chaperonin groEL

Alignment Scores:

Pred. No.:	1.96e-92	Length:	546
Score:	1823.00	Matches:	378
Percent Similarity:	80.51%	Conservative:	64
Best Local Similarity:	68.85%	Mismatches:	99
Query Match:	61.42%	Indels:	8
DB:	2	Gaps:	5

US-09-077-574A-1 (1-1647) x S22347 (1-546)



```
Db        101 GlnGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db        121 AspLeuAlaValAsnGluValValAlaGluLeuLeuLysLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db        141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnAlaGluAlaGluIleGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db        161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db        241 ValGlnThrSerLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db        341 GluIleAspAlaArgValGlyGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuLeu 420

Qy       1261 CGC---TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317

Db        421 ArgAlaSerThrLysIle------ThrAlaLysGlyValAsnAlaAspGlnGluAlaGly 438

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377

Db        439 IleAsnIleValArgArgAlaIleGlnAlaProAlaArgGlnIleThrThrAsnAlaGly 458

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAA---CCAAAAGATGGTTTTGGATTT 1434

Db        459 GluGluAlaSerValIleValGlyLysIleLeuGluAsnThrSerGluThrPheGlyTyr 478

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494

Db        479 AsnThrAlaAsnGlyGluTyrGlyAspLeuIleSerLeuGlyIleValAspProValLys 498

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554

Db        499 ValValArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuIleThrThrGlu 518

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGT 1608

Db        519 AlaMetIleAlaGluLeu------ProLysLysAspAlaAlaProAlaGlyMetProGly 536

Qy       1609 GGTATGGGTGGTATGGGTGGTATGGAC 1635

Db        537 GlyMetGlyGlyMetGlyGlyMetAsp 545
```

Search completed: January 28, 2004, 13:30:02
Job time : 57.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:28:52 ; Search time 60.5 Seconds
(without alignments)
11256.041 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 1554272

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125848_11734/app_query.fasta_1.1799
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09077574_@CGN_1_1_13_@runat_28012004_125848_11734
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1862.5	62.8	547	12	US-10-369-493-17524	Sequence 17524, A
2	1857	62.6	547	12	US-10-369-493-10014	Sequence 10014, A
3	1840.5	62.0	544	12	US-10-369-493-11760	Sequence 11760, A
4	1839.5	62.0	547	12	US-10-369-493-7917	Sequence 7917, Ap
5	1835.5	61.8	542	12	US-10-369-493-10295	Sequence 10295, A
6	1834	61.8	541	12	US-10-369-493-9287	Sequence 9287, Ap
7	1833	61.8	544	12	US-10-369-493-15468	Sequence 15468, A
8	1833	61.8	544	12	US-10-369-493-16217	Sequence 16217, A
9	1828.5	61.6	543	12	US-10-369-493-15265	Sequence 15265, A
10	1826	61.5	542	12	US-10-369-493-11922	Sequence 11922, A
11	1812.5	61.1	551	12	US-10-369-493-12330	Sequence 12330, A
12	1811	61.0	540	12	US-10-369-493-15838	Sequence 15838, A
13	1810	61.0	552	12	US-10-369-493-12170	Sequence 12170, A
14	1786	60.2	545	12	US-10-369-493-167	Sequence 167, App
15	1780.5	60.0	544	12	US-10-369-493-17132	Sequence 17132, A
16	1778.5	59.9	545	12	US-10-228-167A-2	Sequence 2, Appli
17	1771.5	59.7	543	12	US-10-369-493-11916	Sequence 11916, A
18	1770	59.6	536	12	US-10-369-493-4511	Sequence 4511, Ap
19	1770	59.6	540	12	US-10-369-493-7269	Sequence 7269, Ap
20	1746	58.8	545	15	US-10-192-419-2	Sequence 2, Appli
21	1746	58.8	548	11	US-09-415-849-1	Sequence 1, Appli
22	1746	58.8	548	11	US-09-276-455-10	Sequence 10, Appl
23	1743.5	58.7	544	12	US-10-369-493-23094	Sequence 23094, A
24	1737.5	58.5	547	12	US-10-369-493-10862	Sequence 10862, A
25	1721	58.0	521	12	US-10-369-493-9597	Sequence 9597, Ap
26	1720.5	58.0	546	12	US-09-882-227-418	Sequence 418, App
27	1716.5	57.8	540	12	US-10-369-493-181	Sequence 181, App
28	1716.5	57.8	547	15	US-10-046-649-2	Sequence 2, Appli
29	1710	57.6	544	9	US-09-841-132-400	Sequence 400, App
30	1708.5	57.6	548	12	US-10-369-493-513	Sequence 513, App
31	1708	57.5	544	12	US-10-289-762-153	Sequence 153, App
32	1698.5	57.2	542	12	US-10-369-493-17907	Sequence 17907, A
33	1693	57.0	551	12	US-10-369-493-20184	Sequence 20184, A
34	1683.5	56.7	641	15	US-10-267-311-51	Sequence 51, Appl
35	1658	55.9	551	12	US-10-369-493-18994	Sequence 18994, A
36	1650	55.6	544	10	US-09-759-272B-5	Sequence 5, Appli
37	1638.5	55.2	576	12	US-10-369-493-8322	Sequence 8322, Ap
38	1636	55.1	541	15	US-10-156-761-11465	Sequence 11465, A
39	1635	55.1	538	12	US-10-369-493-2950	Sequence 2950, Ap
40	1628.5	54.9	542	15	US-10-156-761-12526	Sequence 12526, A
41	1610.5	54.3	544	12	US-10-369-493-20276	Sequence 20276, A
42	1597	53.8	541	11	US-09-880-505-160	Sequence 160, App
43	1597	53.8	541	12	US-10-205-979-40	Sequence 40, Appl
44	1597	53.8	541	14	US-10-051-643-160	Sequence 160, App
45	1595.5	53.8	540	10	US-09-847-637B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-369-493-17524
; Sequence 17524, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17524
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-17524

Alignment Scores:

Pred. No.:	6.45e-144	Length:	547
Score:	1862.50	Matches:	366
Percent Similarity:	81.39%	Conservative:	80
Best Local Similarity:	66.79%	Mismatches:	97
Query Match:	62.75%	Indels:	5
DB:	12	Gaps:	2

US-09-077-574A-1 (1-1647) x US-10-369-493-17524 (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||||||||:::|||       ||||||||   ::::::       ||||||
Db          1 MetAlaAlaLysGluIleIlePheSerGluLysAlaArgSerArgMetValHisGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||||||||||||||   |||||||||||||||||||||:::||||||:::
Db         21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlySerProIleIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||   |||||||||||||||||||||||||||||:::|||||||||||||   |||
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||  ||||||||||||||||||||||||||||||||||||||||||||:::
Db         81 ThrAsnAspHisAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||:::|||          |||||       |||:::|||||||||   |||
Db        121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||   ||||||||||||   ||||||||||||||||||||   :::||||||||||
Db        141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||   ||||||||||||   |||||||||||:::||||||
Db        161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCCGTGGCTACCTC 600
              ||||||      ||||||||||||||||||||:::|||||||||||||||
Db        181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||:::   |||  :::   :::   ||||||||||||||||||||
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   ::::::||||||||:::|||::::::|||:::|||   :::|||:::   |||
Db        221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||   ::::::|||||||||||||:::||||||::||||||||||||||||||||
Db        241 AlaLysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   ::::::|||   ||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||:::||:::||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::   |||   ||||||   |||
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   :::   ||||||||||||||||||:::||||||   |||   ::::::
Db        321 LysLysValArgValSerLysGluAsnThrThrIleIleAspGlyIleGlyAspAsnAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||      ||||||||||||||:::   ||||||||||||||||:::||||||||
Db        341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||::||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||||   ||||||:::||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||:::::|||||||||||   |||   :::
Db        401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::|||       :::   ::::::|||   |||::::::|||:::      |||:::
Db        421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 |||      |||::::::|||   |||||||||:::|||   ||||||||||
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::::::::   ||||||:::|||   ||||||   ||||||:::||||||
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::||||||:::   |||::::::      |||::::::||||||   ||||||||
Db        481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   |||||||||||||||||||:::|||   |||::::::|||||||||
Db        501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCT---------GGCGGTGGT 1611
              :::||||||            ||||||||||||   |||   |||          |||||||||
Db        521 ValAlaGlu------AlaProLysLysAspGluProThrProProAlaAlaGlyGlyGly 538

Qy       1612 ATGGGTGGTATGGGTGGTATGGAC 1635
              ||||||||||||||||||||||||
Db        539 MetGlyGlyMetGlyGlyMetAsp 546
```

RESULT 2
US-10-369-493-10014
; Sequence 10014, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10014
; LENGTH: 547
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-10014

Alignment Scores:

Pred. No.:	1.82e-143	Length:	547
Score:	1857.00	Matches:	382
Percent Similarity:	82.55%	Conservative:	72
Best Local Similarity:	69.45%	Mismatches:	90
Query Match:	62.57%	Indels:	6
DB:	12	Gaps:	4

US-09-077-574A-1 (1-1647) x US-10-369-493-10014 (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::||||||:::   |||       ||||||   |||:::      ||||||
Db          1 MetAlaAlaLysGluValLysPheGlyGluAlaAlaArgAlaLysMetLeuAsnGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||:::|||:::|||   :::||||||||||||||||||||||||||||||||
Db         41 AspLysSerTrpGlyAlaProArgMetThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||:::||||||:::   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValSerSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrAlaAspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||:::|||   ||||||||||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyMetLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||   ||||||   ||||||        |||      |||::::::          :::
Db        121 AspLeuAlaValGluAlaValValValGlyLeuLysGlyIleSerArgGluValAlaAsn 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                 :::||||||||||||||||||   ||||||||||||||||||      :::|||:::
Db        141 SerGlnGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspLysValValGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::||||||||||||||   ||||||||||   ||||||||||||||||||||||||||
Db        161 MetIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        181 LysGlyLeuGluThrThrLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||   :::||||||:::   :::::::::||||||   |||
Db        201 SerProTyrPheValThrAsnAlaAspLysMetLeuValGlnMetGluAsnProLeuIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||       ||||||||||||:::::::::::   :::|||   |||||||||
Db        221 LeuLeuValGluLysLysIleSerAsnLeuGlnGlnIleLeuGlnIleLeuGluGlyAla 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::   :::||||||||:::||||||||||||||||||||||||||||||||||||
Db        241 ValGlnSerSerArgProLeuMetIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||   |||:::|||   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||:::|||||||||||   ||||||||||
Db        281 PheGlyAspArgArgLysAlaMetMetGluAspIleAlaThrLeuThrGlyGlyValLeu 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              :::   ||||||   |||:::||||||||||||||:::::::      ||||||   |||
Db        301 ValSerGluAspValGlyValLysLeuGluAsnValThrMetAspMetLeuGlyMetAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||   :::|||:::   |||||:::|||||||||:::||||||||||||      |||
Db        321 LysSerIleValValThrLysGluAspThrThrIleIleAspGlyAlaGlyAspHisGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 ||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        341 AlaIleLysAlaArgValAsnGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   |||||||||   |||:::||||||:::|||||||||||:::|||||||||
Db        381 ValGlyGlyAlaThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||||||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTC---CTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
              |||:::   :::      |||   :::::::::   |||:::   |||:::      |||
Db        421 ArgAlaArgGluAlaSerLeuThrAsnLeuGlnGlyAlaAsnHisAspGlnGlnValGly 440

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
              :::||||||:::   |||:::|||||||||||||||||   ||||||:::||||||||
Db        441 IleAsnIleValThrArgAlaLeuGluGluProLeuArgIleIleAlaSerAsnAlaGly 460

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTT 1434
                 |||||||||:::||||||   :::|||   |||   |||:::      |||||||||
Db        461 AlaGluGlySerValValValAsnArgValValGluThrLysGluThrAsnPheGlyPhe 480

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
              |||||||||:::|||   ||||||||||:::   :::|||||||||||||   |||
Db        481 AsnAlaAlaThrGlyValTyrGluAspLeuValAlaSerGlyValIleAspProAlaLys 500

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
              |||   |||   |||||||||   ||||||||||||||   |||::::::||||||||
Db        501 ValValArgHisAlaLeuGlnAlaAlaAlaSerValAlaGlyLeuMetIleThrThrGlu 520

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAAGATATGCCT---ATGCCTGGCGGTGGT 1611
                  :::||||||   |||         ||||||   |||   ||||||||||||
Db        521 AlaMetValAlaGluLeuPro---------LysAspGluProAlaMetProGlyGlyAsp 537

Qy       1612 ATGGGTGGTATGGGTGGTATGGACGGTATG 1641
              ||||||||||||||||||||   ||||||
Db        538 MetGlyGlyMetGlyGlyMetGlyGlyMet 547
```

RESULT 3
US-10-369-493-11760
; Sequence 11760, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11760
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11760

Alignment Scores:

Pred. No.:	4.09e-142	Length:	544
Score:	1840.50	Matches:	376
Percent Similarity:	81.95%	Conservative:	69
Best Local Similarity:	69.24%	Mismatches:	95
Query Match:	62.01%	Indels:	3
DB:	12	Gaps:	3

US-09-077-574A-1 (1-1647) x US-10-369-493-11760 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::||||||:::   |||          ||||||||||||:::   :::||||||
Db          1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuLysGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||   ||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||||||||:::|||  ||||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              |||||||||||||||||||||||||||||||:::|||:::||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||   |||||       |||   |||:::|||              |||
Db        121 AspLeuAlaValAlaGluValValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              :::|||:::|||||||||||||||||||||||||||   :::      ||||||
Db        141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              |||||||||||||||   :::||||||      ||||||||||||||||||||||||||||
Db        161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      ||||||   |||:::|||||||||||||||:::|||||||||||||||||||
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              |||||||||||||||||||||||||||||||   :::|||::::::   ||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||||:::::::::::::   |||||||||:::||||||   |||
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              :::      :::||||||:::||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnThrGlyLysProLeuValIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||   |||::::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   |||||||||||||||:::|||:::|||       ||||||   :::
Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   |||   |||||||||||||||||||||||||||||||||:::
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||:::   ||||||   ||||||:::|||||||||||||||||::|||||||||
Db        341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   :::||||||   |||:::|||||||||||||||||::::::|||||||||
Db        381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              |||||||||||||||||||:::|||||||||||||||||||||||||   |||   :::
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||          :::      :::|||      :::|||||:::   ||||||:::
Db        421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspAspGlnGluAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              |||||:::|||::::::|||:::     :::|||||||||||   |||||||||
Db        440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCA---AAAGATGGTTTTGGATTTAAT 1437
              |||   ||||||||||||   |||:::   :::      :::|||   :::|||:::|||
Db        460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
              |||   :::||||||||||   |||||||||      |||:::::|||||   ||||||
Db        480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
              |||   |||||||||||||||||||||||||||||||||||||:::|||||||||
Db        500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
              |||||||||   |||:::   :::      ||||||   ||||||||||||||||||
Db        520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet 539

Qy       1615 GGTGGTATG 1623
              |||||||||
Db        540 GlyGlyMet 542
```

RESULT 4
US-10-369-493-7917
; Sequence 7917, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7917
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-7917

Alignment Scores:

Pred. No.:	4.95e-142	Length:	547
Score:	1839.50	Matches:	372
Percent Similarity:	80.66%	Conservative:	70
Best Local Similarity:	67.88%	Mismatches:	103
Query Match:	61.98%	Indels:	3
DB:	12	Gaps:	2

US-09-077-574A-1 (1-1647) x US-10-369-493-7917 (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheAspThrAspAlaArgAspArgMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuSerAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspGluAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 LysGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaThrSerLysValValGluAlaIleLysAlaAlaAlaArgProValAsnAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerHisGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluAlaGlnIleGlyArg 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 PheIleAlaAspAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAsn 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysGlyLeuGluThrGluValGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetThrAlaGluLeuAspAspValTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerSerLeuGlnProMetValProLeuLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 IleGlnSerGlnLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyMetLysLeuGluAsnValThrIleAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysIleSerIleAsnLysAspAsnThrThrIleValAspGlyAsnGlyAspLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleAspAlaArgValAlaGlnIleArgAsnGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyMetThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValValGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 GlnGlyGlyLysAlaLeuAspGlyLeuThrGlyGluAsnProAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 ThrIleValArgArgAlaLeuGluAlaProLeuArgGlnIleAlaGlnAsnAlaGlyVal 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
Db        461 AspGlySerValValAlaGlyLysValArgGluSerAsnGluLysSerPheGlyPheAsn 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        481 AlaGlnThrGluGluTyrGlyAspMetPheLysPheGlyValIleAspProAlaLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        501 ValArgThrAlaLeuGluAspAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db        521 MetIleAlaAspLysProGluProLysSer------ProAlaGlyGlyProGlyMetGly 538

Qy       1618 GGTATGGGTGGTATGGACGGTATG 1641
Db        539 GlyMetGlyGlyMetAspGlyMet 546
```

RESULT 5
US-10-369-493-10295
; Sequence 10295, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10295
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
US-10-369-493-10295

```
Alignment Scores:
Pred. No.:              1.05e-141      Length:          542
Score:                  1835.50        Matches:         365
Percent Similarity:     81.85%         Conservative:    77
Best Local Similarity:  67.59%         Mismatches:      97
Query Match:            61.84%         Indels:          1
DB:                     12             Gaps:            1

US-09-077-574A-1 (1-1647) x US-10-369-493-10295 (1-542)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||:::|||||||||||||     |||||||||||||||   :::|||||||||
Db          2 AlaLysGlnIleLeuPheAspArgAspAlaArgGluLysLeuLysLysGlyValAspAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::|||||||||||||||||||||||||||||||||||||:::|||||||||||
Db         22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIleLeuAspLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
                 ||||||||||||  |||||||||||||||||||||||||||||:::|||||||||
Db         42 LysPheGlySerProAlaIleThrLysAspGlyValSerValAlaLysAspIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              :::|||      |||||||||||||||||:::|||||||||||||||   ||||||:::
Db         62 LysAspAlaIleGluAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   |||||||||||||||||||||||||||||||   ||||||||:::
Db         82 AspGlnAlaGlyAspGlyThrThrThrAlaThrValLeuThrGlnAlaIlePheAsnAla 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||:::|||   ||||||||||||   |||||||||   :::|||||||||||||||||||
Db        102 GlyLeuLysAsnValAlaAlaGlyAlaAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||   |||:::      :::|||          :::|||          :::
Db        122 AlaValSerAlaIleIleAlaAspLeuLysValGlnSerLysLysIleSerAsnSerAsn 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||||||||||   ||||||||||||||:::|||      ||||||   :::|||
Db        142 GluIleAlaGlnValAlaThrIleSerAlaAsnAsnAspHisGluIleGlyLysMetIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||   ||||||   |||||||||||||   ||||||||||||||||||||||||||||||
Db        162 AlaHisAlaMetAspLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
                ||||||   :::      ||||||||||||:::|||||||||||||||||||||||
Db        182 ThrGluThrGluValLysThrValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||   :::||||||      ||||||||   ||||||||||||
Db        202 TyrPheValThrAsnThrAspLysMetGluValGluLeuAspArgProTyrIleLeuIle 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::||||||:::::|||||||||:::::|||||||||||||||||||   :::
Db        222 TyrAspLysLysValSerSerMetLysGluLeuLeuProIleLeuGluGlnValValGln 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                  ||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        242 SerGlyArgProLeuLeuIleIleAlaGluAspValAspGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||:::|||||||||||||::::::   ||||||||||||||||||||||
Db        262 ValValAsnLysIleArgGlyAlaLeuLysIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        282 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||   ||||||||||||   ::|||      ||||||||||||::::::
Db        302 GluGluSerGlyTyrLysLeuGluAsnAlaThrLeuGluTyrLeuGlyThrAlaGluLys 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::   |||||||||:::|||||||||:::|||:::|||:::|||      ::::::|||
Db        322 IleAsnIleAspLysAspAsnThrThrValValAsnGlySerGlyLeuLysAspAsnIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
                 |||||||||   |||||:::::::|||:::|||   |||:::|||||||||||||||
Db        342 ValAlaArgValAsnGlnIleLysSerGlnMetGluAsnThrThrSerAspTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              |||||||||||||||||||||||||||||||   |||||||||||:::::::::::|||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaIleMetTyrIleGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||||   |||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaThrGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuHisAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||||||||||||||   |||   :::|||:::
Db        402 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                    ||||||::::::         ::::::||||||      |||:::   |||
Db        422 GlyAlaAlaLeuAspAsnValAlaPheHisAsnGluAspGluLysThrGlyIleGlnIle 441

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              ||||||   ::::::|||   |||||||||   |||      |||||||||   ||||||
Db        442 IleArgThrAlaIleGluSerProLeuArgAlaIleValTyrAsnAlaGlyLeuGluGly 461

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||::::::|||:::||||||:::|||           :::|||:::||||||
Db        462 SerValIleValGlnLysValLysGluGlyThrGlyAspTyrGlyTyrAsnAlaArgGlu 481

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
                    ||||||   :::|||   |||||:::|||||||||   ||||||||||:::
Db        482 AspArgTyrGluAlaMetIleAlaAlaGlyIleIleAspProThrLysValThrArgLeu 501

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||:::|||||||||||||||||||||||||||||||||||||||   :::|||
Db        502 AlaLeuGluAsnAlaAlaSerValAlaSerLeuLeuLeuThrThrGluCysValValAla 521

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATG---CCTATGCCTGGCGGTGGTATGGGTGGTATG 1623
              ::::::|||||   |||      |||   ||||||   ||||||||||||||||||||
Db        522 AspGluProGluGluLysSerAlaMetProProMetGlyGlyGlyGlyMetGlyGlyMet 541

RESULT 6
US-10-369-493-9287
; Sequence 9287, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9287
;   LENGTH: 541
;   TYPE: PRT
;   ORGANISM: Xylella fastidiosa
US-10-369-493-9287

Alignment Scores:
Pred. No.:              1.39e-141      Length:          541
Score:                  1834.00        Matches:         360
Percent Similarity:     81.03%         Conservative:    80
```

```
Best Local Similarity:  66.30%          Mismatches:    101
Query Match:            61.79%          Indels:          2
DB:                     12              Gaps:            1

US-09-077-574A-1 (1-1647) x US-10-369-493-9287 (1-541)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||||||||:::|||      |||||||||   ::::::      ||||||
Db          1 MetAlaAlaLysGluIleIlePheSerGluLysAlaArgSerArgMetValTyrGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||||||||||||||   |||||||||||||||||||||:::||||||:::
Db         21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlySerProIleIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||   ||||||||||||||||||||||||||:::||||||||||||   |||
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||   |||||||||||||||||||||||||||||||||||||||||||:::
Db         81 ThrAsnAspHisAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||:::|||          ||||||          |||:::|||||||||   |||
Db        121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                 |||   ||||||||||||   ||||||||||||||||||||   :::||||||||
Db        141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||   ||||||||||||   |||||||||||||:::||||||
Db        161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
                    ||||||      ||||||||||||||||||||||:::||||||||||||||
Db        181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::   |||   :::      :::   ||||||||||||||||||||
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   ::::::||||||||:::|||::::::|||:::|||   :::|||:::   |||
Db        221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||   ::::::||||||||||||:::||||||:::||||||||||||||||||||
Db        241 AlaLysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   ::::::|||   |||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||||||:::|||:::||||||||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::   |||   ||||||   |||
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   :::   ||||||||||||||||||:::||||||   |||   ::::::
Db        321 LysLysValArgValSerLysGluAsnThrThrIleIleAspGlyIleGlyAspAsnAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||      ||||||||||||||:::   ||||||||||||||:::||||||||||
Db        341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||:::||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||   ||||||||||||||||||   ||||||:::||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||:::::||||||||||   |||   :::
Db        401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::|||      :::   ::::::|||   |||::::::|||:::      |||:::
Db        421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 |||      |||::::::|||   |||||||||:::|||   ||||||||||
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::::::::   ||||||:::|||   ||||||   ||||||:::||||||
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::||||||:::   |||::::::|||   |||::::::||||||   ||||||
Db        481 AlaThrGlyGluPheGlyAspMetValLysLeuGlyIleLeuAspProThrLysValAla 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   ||||||||||||||||||||:::|||   |||::::::||||||||||
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              :::||||||          ||||||||||||   |||   |||                ||||||
Db        521 ValAlaGlu------AlaProLysLysAspGluProThrProProAlaAlaGlyGlyGly 538

Qy       1621 ATGGGTGGT 1629
              |||||||||
Db        539 MetGlyGly 541
```

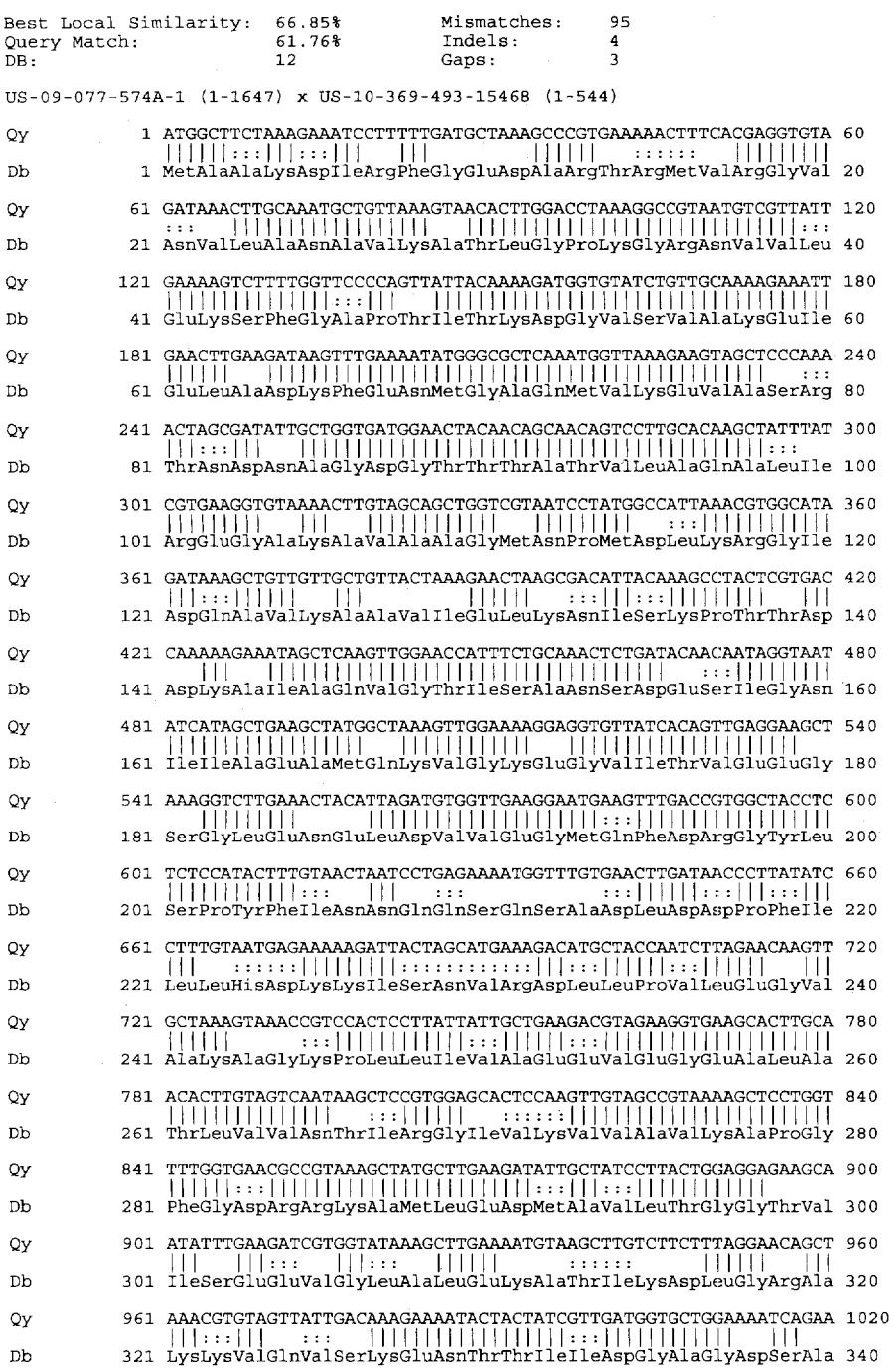
RESULT 7
US-10-369-493-15468
; Sequence 15468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15468
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15468

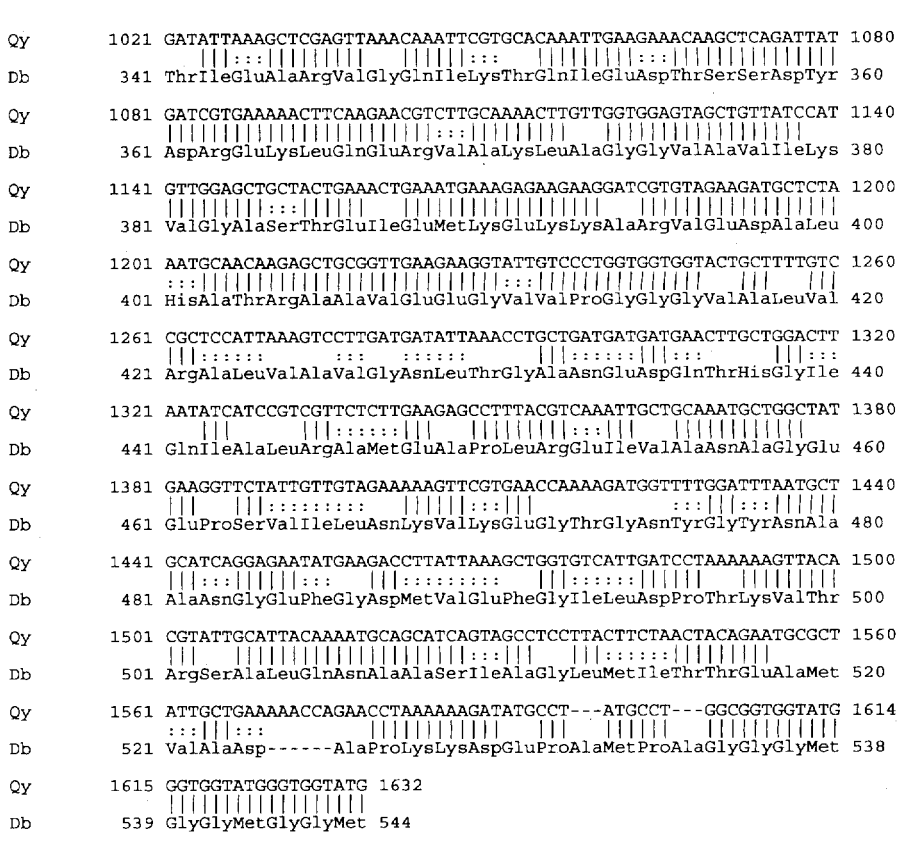
Alignment Scores:

Pred. No.:	1.68e-141	Length:	544
Score:	1833.00	Matches:	365
Percent Similarity:	81.87%	Conservative:	82

Best Local Similarity: 66.85% Mismatches: 95
Query Match: 61.76% Indels: 4
DB: 12 Gaps: 3

US-09-077-574A-1 (1-1647) x US-10-369-493-15468 (1-544)





RESULT 8
US-10-369-493-16217
; Sequence 16217, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16217
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-16217

Alignment Scores:
Pred. No.: 1.68e-141 Length: 544
Score: 1833.00 Matches: 365
Percent Similarity: 81.87% Conservative: 82

```
Best Local Similarity:  66.85%        Mismatches:    95
Query Match:            61.76%        Indels:         4
DB:                     12            Gaps:           3

US-09-077-574A-1 (1-1647) x US-10-369-493-16217 (1-544)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspIleArgPheGlyGluAspAlaArgThrArgMetValArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspGlnAlaValLysAlaAlaValIleGluLeuLysAsnIleSerLysProThrThrAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 AspLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetGlnLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 SerGlyLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnSerAlaAspLeuAspAspProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisAspLysLysIleSerAsnValArgAspLeuLeuProValLeuGluGlyVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnThrIleArgGlyIleValLysValValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluValGlyLeuAlaLeuGluLysAlaThrIleLysAspLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValGlnValSerLysGluAsnThrThrIleIleAspGlyAlaGlyAspSerAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 ThrIleGluAlaArgValGlyGlnIleLysThrGlnIleGluAspThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaSerThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaLeuValAlaValGlyAsnLeuThrGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyThrGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaAsnGlyGluPheGlyAspMetValGluPheGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCT---GGCGGTGGTATG 1614
Db        521 ValAlaAsp------AlaProLysLysAspGluProAlaMetProAlaGlyGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATG 1632
Db        539 GlyGlyMetGlyGlyMet 544
```

RESULT 9
US-10-369-493-15265
; Sequence 15265, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15265
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15265

```
Alignment Scores:
Pred. No.:              3.93e-141     Length:          543
Score:                  1828.50       Matches:         374
Percent Similarity:     81.77%        Conservative:    70
```

```
Best Local Similarity:  68.88%        Mismatches:    96
Query Match:            61.61%        Indels:         3
DB:                     12            Gaps:           3

US-09-077-574A-1 (1-1647) x US-10-369-493-15265 (1-543)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||||||:::          ||   ||||||||||||:::   :::||||||
Db          1 MetAlaAlaLysGluValLysProGlyAlaSerAlaArgGluLysMetLeuLysGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||   |||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||:::|||   ||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlyAlaProProIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              |||||||||||||||||||||||||||||||:::|||:::|||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   ||||||||   :::|||||||||||||
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||   ||||||       |||   |||:::|||              |||
Db        121 AspLeuAlaValAlaGluValValLysAspLeuGlnAlaLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                 :::|||:::|||||||||||||||||||||||||||||   :::       |||||
Db        141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluArgGlnIleGlyLeu 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
                 ||||||||||||||   :::||||||      ||||||||||||||||||||||||
Db        161 AspIleAlaGluAlaMetGlnArgValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      ||||||   |||:::|||||||||||||||:::|||||||||||||||||||
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              |||||||||||||||||||||||||||||||||   :::|||::::::   ||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::||||||||:::::::::::::::   |||||||||:::||||||   |||
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::      :::||||||:::|||||||||||||||||||||||||||||   ||||||
Db        241 ValGlnThrGlyLysProLeuValIleIleAlaGluAspValGluGlyGluArgLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              |||||||||||||||||||||||||   |||::::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   |||||||||||||||:::|||:::|||      ||||||   :::
Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyLysSer 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   |||   |||||||||||||||||||||||||||||:::
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||:::   ||||||   ||||||:::|||||||||||||||||||:::|||||||||
Db        341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   :::||||||   |||:::|||||||||||||||||:::::|||||||||
Db        381 ValGlyGlySerThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              |||||||||||||||||||:::|||||||||||||||||||||||||   |||   :::
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||         :::      :::|||      :::||||||:::   ||||||:::
Db        421 ArgSerSerThrLysIle---ThrValLysGlyValAsnAspAspGlnGluAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||:::|||:::::|||:::      :::||||||||||   |||||||||
Db        440 AsnIleValArgLysAlaLeuGlnSerLeuValArgGlnIleAlaGluAsnAlaGlyAsp 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCA---AAAGATGGTTTTGGATTTAAT 1437
              |||   |||||||||||   |||:::   :::      :::|||   :::|||:::|||
Db        460 GluAlaSerIleValValGlyLysIleLeuAspLysAsnGluAspAsnTyrGlyTyrAsn 479

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
              |||   :::|||||||||   |||||||||      |||:::::|||||   ||||||
Db        480 AlaGlnThrGlyGluTyrGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
                 |||   |||||||||||||||||||||||||||||:::|||:::||||||||||
Db        500 ValArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerValLeuIleThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
              |||||||||   |||:::   :::      ||||||   ||||||||||||||||||
Db        520 MetIleAlaGluLeuProLysLysGluSerAlaMetProGlnMetProGlyGlyGlyMet 539

Qy       1615 GGTGGTATG 1623
              |||||||||
Db        540 GlyGlyMet 542

RESULT 10
US-10-369-493-11922
; Sequence 11922, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11922
;   LENGTH: 542
;   TYPE: PRT
;   ORGANISM: Mesorhizobium loti
US-10-369-493-11922

Alignment Scores:
Pred. No.:              6.3e-141      Length:         542
Score:                  1826.00       Matches:        366
Percent Similarity:     82.59%        Conservative:   80
```

```
Best Local Similarity:  67.78%          Mismatches:     92
Query Match:            61.52%          Indels:          2
DB:                     12              Gaps:            2

US-09-077-574A-1 (1-1647) x US-10-369-493-11922 (1-542)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::||||||:::   |||   :::   |||||||||||:::   |||||||||
Db          1 MetAlaAlaLysGluValLysPheHisSerAspAlaArgGluLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                 :::   ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||:::|||   |||||||||||||||||:::|||||||||||||||
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              |||||||||||||||||||||||||||||||||||||||||:::|||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrSerAspLeuAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              :::||||||   |||   |||||:::|||   |||||||||   :::|||||||||||||
Db        101 LysGluGlyAlaLysAlaValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||:::   :::||||||          :::
Db        121 AspLysAlaValGluAlaIleValGlnGluLeuLysThrAsnAlaArgLysValThrArg 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                  ||||||||||||||||||||||||||||||||   |||      ||||||
Db        141 AsnAspGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAspAlaGluIleGlyArg 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
                 :::|||||||||||   |||||||||      ||||||||||||||||||||||||
Db        161 PheLeuAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      ||||||   |||:::|||||||||||||:::|||||||||||||||||||
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::||||||   :::||||||      ||||||:::   ||||||:::
Db        201 SerProTyrPheIleThrAsnGlnAspLysMetArgValGluLeuGluGluProTyrVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||||:::::::::::::::   :::|||||:::||||||   |||
Db        221 LeuIleHisGluLysLysLeuSerAsnLeuGlnAlaLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::   ::::::|||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnSerSerLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||   |||:::|||   |||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||   |||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrAla 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   |||||||||||||||||||:::|||      ||||||   |||
Db        301 IleSerGluAspLeuGlyIleLysLeuGluAsnValThrLeuGluMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||||||||:::|||||||||||||||||||||||||||||||      :::
Db        321 LysLysValValIleGluLysGluAsnThrThrIleValAspGlyAlaGlySerLysAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              :::|||:::   ||||||   ||||||::::::|||||||||||||||:::|||||||||
Db        341 GluIleGlnGlyArgValSerGlnIleLysSerGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||:::|||||||||||||||||||||||||||||||   |||||||||||||||||
Db        361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   :::||||||   |||:::||||||:::||||||||||||:::||||||:::
Db        381 ValGlyGlySerThrGluValGluValLysGluArgLysAspArgValAspAspAlaMet 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||:::::|||||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValLeuProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::   |||   ||||||   ::::::      ::::::|||:::      |||:::
Db        421 ArgAlaAlaLysAlaLeuAspSerValGlnAlaGluAsnGluAspGlnLysHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 |||:::|||||:::::|||   |||:::||||||||||||   |||||||||
Db        441 GluIleValArgArgAlaIleGluAlaProValArgGlnIleAlaGluAsnAlaGlyAla 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAA---CCAAAAGATGGTTTTGGATTTAAT 1437
              ||||||||||||:::|||   |||:::||||||   |||:::   |||:::|||:::|||
Db        461 GluGlySerIleIleValGlyLysLeuArgGluLysProGluPheGlyTrpGlyTrpAsn 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
              |||   :::      :::   ||||||          ||||||||||||||   ||||||
Db        481 AlaGlnThrAsnAlaPheGlyAspLeuTyrAsnGluGlyValIleAspProValLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
              |||   |||||||||:::|||||||||||||||   ||||||:::||||||||||
Db        501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAAGATATGCCT---ATGCCTGGCGGTGGTATG 1614
              :::|||||||||||:::   :::      :::|||   ||||||   |||||||||
Db        521 MetValAlaGluLysProLysLysGluProAlaValProAlaMetProAlaGlyGlyMet 540
```

```
RESULT 11
US-10-369-493-12330
; Sequence 12330, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12330
;   LENGTH: 551
;   TYPE: PRT
;   ORGANISM: Mesorhizobium loti
US-10-369-493-12330

Alignment Scores:
Pred. No.:              8.08e-140       Length:         551
Score:                  1812.50         Matches:        375
Percent Similarity:     80.65%          Conservative:   71
Best Local Similarity:  67.81%          Mismatches:     96
Query Match:            61.07%          Indels:         11
DB:                     12              Gaps:           5
```

US-09-077-574A-1 (1-1647) x US-10-369-493-12330 (1-551)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 GlnGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaValSerGluValValAlaAlaLeuGlyLysAlaAlaLysLysIleLysThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyAspGluSerValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnThrSerLysProLeuLeuIleIleSerGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyIleLysLeuGluAsnValGlyLeuAsnMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyLysLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleGlnGlyArgValAlaGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCT------GCTGATGATGATGAACTTGCT 1314
Db        421 ArgAlaSerAla---------AsnIleLysAlaAlaGlyAlaAsnAlaAspGlnAlaAla 437

Qy       1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCT 1374
Db        438 GlyIleAsnIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaSerAsnAla 457

Qy       1375 GGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGT---TTTGGA 1431
Db        458 GlyAlaGluAlaSerIleValAlaGlyLysIleLeuGluAsnLysGlyAlaThrPheGly 477

Qy       1432 TTTAATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAA 1491
Db        478 TyrAsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleValAspProVal 497

Qy       1492 AAAGTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACA 1551
Db        498 LysValValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThr 517

Qy       1552 GAATGCGCTATTGCTGAAAAACCAGAACCTAAAAAA---------GATATGCCT------ 1596
Db        518 GluAlaMetIleAlaGluAlaProLysLysGluSerAlaGlyGlyGlyMetProGlyGly 537

Qy       1597 ATGCCTGGCGGTGGTATGGGTGGTATGGGTGGTATGGAC 1635
Db        538 MetGlyGlyGlyGlyMetGlyGlyMetGlyGlyMetAsp 550
```

RESULT 12
US-10-369-493-15838
; Sequence 15838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15838
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15838

Alignment Scores:

Pred. No.:	1.07e-139	Length:	540
Score:	1811.00	Matches:	361
Percent Similarity:	81.73%	Conservative:	82
Best Local Similarity:	66.61%	Mismatches:	95
Query Match:	61.02%	Indels:	4
DB:	12	Gaps:	3

US-09-077-574A-1 (1-1647) x US-10-369-493-15838 (1-540)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspIleArgPheGlyGluAspAlaArgThrArgMetValArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspGlnAlaValLysAlaAlaValIleGluLeuLysAsnIleSerLysProThrThrAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 AspLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetGlnLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 SerGlyLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnSerAlaAspLeuAspAspProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisAspLysLysIleSerAsnValArgAspLeuLeuProValLeuGluGlyVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnThrIleArgGlyIleValLysValValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluValGlyLeuAlaLeuGluLysAlaThrIleLysAspLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValGlnValSerLysGluAsnThrThrIleIleAspGlyAlaGlyAspSerAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 ThrIleGluAlaArgValGlyGlnIleLysThrGlnIleGluAspThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaSerThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaLeuValAlaValGlyAsnLeuThrGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyThrGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaAsnGlyGluPheGlyAspMetValGluPheGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCT---ATGCCT---GGCGGTGGTATG 1614
Db        521 ValAlaAsp------AlaProLysLysAspGluProAlaMetProAlaGlyGlyGlyMet 538

Qy       1615 GGTGGT 1620
Db        539 GlyGly 540
```

RESULT 13
US-10-369-493-12170
; Sequence 12170, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12170
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12170

Alignment Scores:

Pred. No.:	1.3e-139	Length:	552
Score:	1810.00	Matches:	373
Percent Similarity:	80.62%	Conservative:	72
Best Local Similarity:	67.57%	Mismatches:	99
Query Match:	60.98%	Indels:	8
DB:	12	Gaps:	4

US-09-077-574A-1 (1-1647) x US-10-369-493-12170 (1-552)

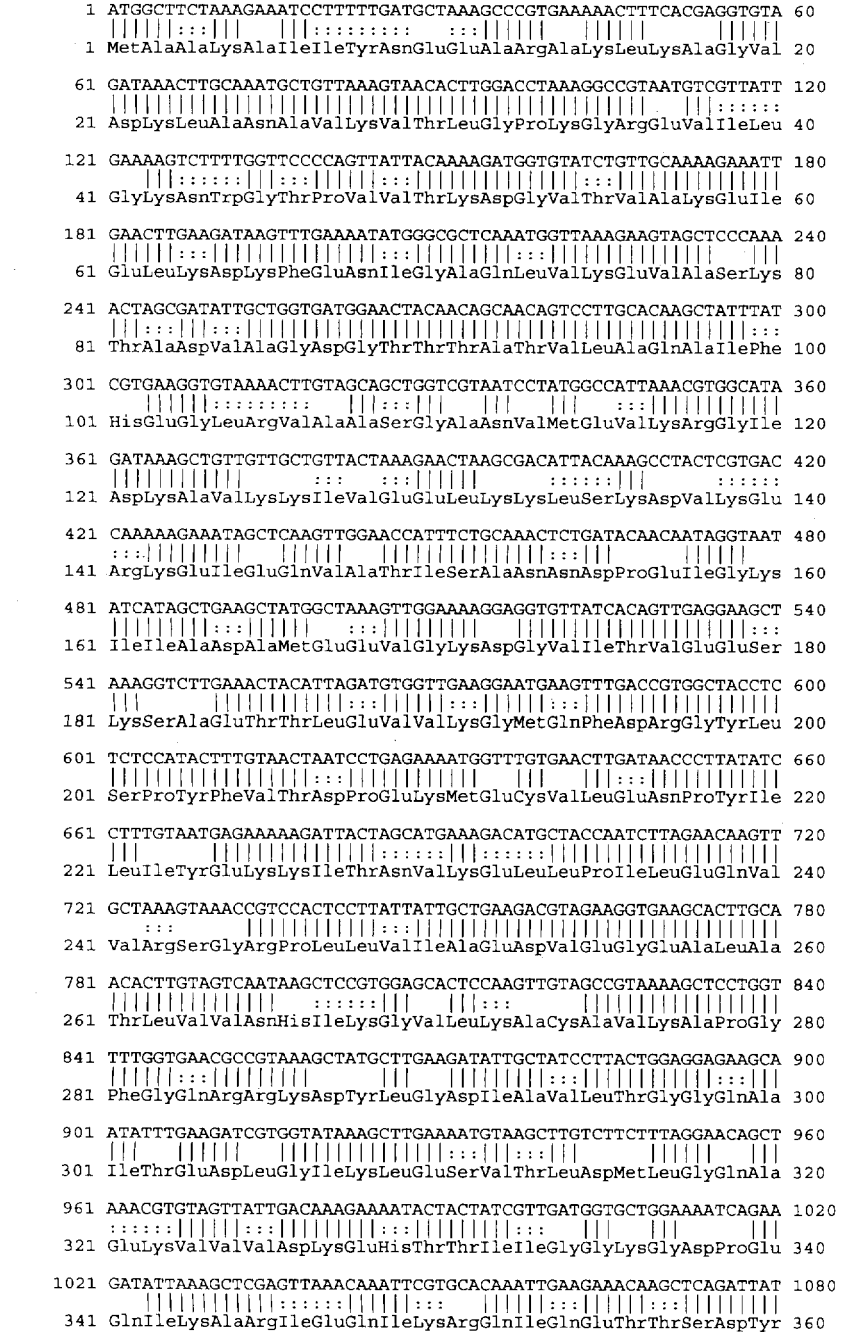
```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 GlnGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaValSerAspValValTrpThrLeuIleLysAsnAlaThrLysIleLysThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerGluGluValAlaGlnValGlyThrIleAlaGlyAsnGlyAspGluSerValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnThrSerLysProLeuLeuIleIleSerGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyIleLysLeuGluAsnValGlyLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyLysLysGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleGlnGlyArgValAlaGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaSerLeuSerIleAsnAlaVal---GlyAlaAsnSerAspGlnThrAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        440 SerIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaAlaAsnAlaGlyAla 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGT---TTTGGATTTAAT 1437
Db        460 GluAlaSerIleValAlaGlyLysIleLeuGluAsnLysGlyAlaThrPheGlyPheAsn 479

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        480 AlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        500 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAA----------GATATGCCT------ATG 1599
Db        520 MetIleAlaGluAlaProLysLysGluSerAlaGlyGlyGlyGlyMetProGlyGlyMet 539

Qy       1600 CCTGGCGGTGGTATGGGTGGTATGGGTGGTATGGAC 1635
Db        540 GlyGlyGlyGlyMetGlyGlyMetGlyGlyMetAsp 551
```

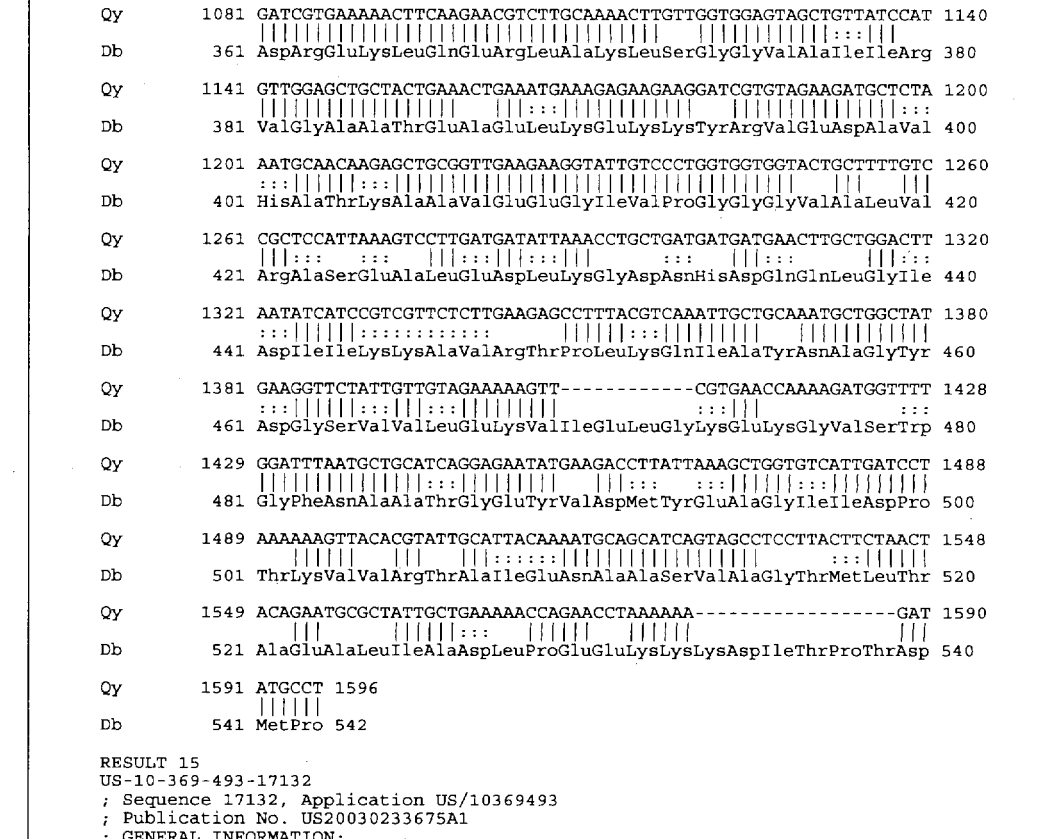
RESULT 14
US-10-369-493-167
; Sequence 167, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 167
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Aquifex aeolicus
US-10-369-493-167

Alignment Scores:

Pred. No.:	1.19e-137	Length:	545
Score:	1786.00	Matches:	352
Percent Similarity:	82.29%	Conservative:	94
Best Local Similarity:	64.94%	Mismatches:	86
Query Match:	60.18%	Indels:	10
DB:	12	Gaps:	2

US-09-077-574A-1 (1-1647) x US-10-369-493-167 (1-545)





RESULT 15
US-10-369-493-17132
; Sequence 17132, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17132
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17132

Alignment Scores:

Pred. No.:	3.37e-137	Length:	544
Score:	1780.50	Matches:	355
Percent Similarity:	81.02%	Conservative:	89
Best Local Similarity:	64.78%	Mismatches:	95
Query Match:	59.99%	Indels:	9
DB:	12	Gaps:	4

US-09-077-574A-1 (1-1647) x US-10-369-493-17132 (1-544)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db          2 AlaLysAspIleLysPheSerGluAspAlaArgArgSerMetLeuArgGlyValAspLys 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db         22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db         42 LysPheGlySerProLeuIleThrAsnAspGlyValThrIleAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db         62 GluAspAlaPheGluAsnMetGlyAlaLysLeuValAlaGluValAlaSerLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db         82 AspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaMetIleArgGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        102 GlyLeuLysAsnValThrSerGlyAlaAsnProMetValIleArgLysGlyIleGluLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        122 AlaThrGlnValAlaValGluGluLeuSerLysIleSerLysProIleGluGlyLysAsp 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        142 SerIleAlaGlnValAlaAlaIleSer---SerAlaAspAspGluValGlyLysIleIle 160

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        161 AlaGluAlaMetGluArgValGlyAsnAspGlyValIleThrIleGluGluSerLysGly 180

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db        181 PheSerThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrAlaSerPro 200

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        201 TyrMetValThrAspSerAspLysMetGluAlaValLeuAspAsnProTyrValLeuIle 220

Qy        667 AATGAGAAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        221 ThrAspLysLysIleSerAsnIleGlnGluValLeuProValLeuGluGlnValValGln 240

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        241 GlnGlyLysProIleLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 260

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db        261 ValValAsnLysLeuArgGlyThrPheAsnAlaValAlaValLysAlaProGlyPheGly 280

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db        281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleThr 300

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db        301 GluAspLeuGlyLeuAspLeuLysSerAlaAsnIleThrGlnLeuGlyArgAlaSerLys 320

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db        321 ValValValThrLysGluAsnThrThrIleValGluGlyAlaGlyGluSerAspLysIle 340

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db        341 AlaAlaArgValAsnGlnIleLysAlaGlnIleGluGluThrThrSerAspPheAspLys 360

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db        361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValLeuLysValGly 380

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db        381 AlaAlaThrGluThrGluMetLysGluArgLysLeuArgIleGluAspAlaLeuAsnSer 400

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db        401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuValAsnVal 420

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db        421 IleLysAlaValSerSerIle---GlyAlaGluGlyAspGluAlaThrGlyValAsnIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db        440 ValLeuArgAlaLeuGluGluProValArgGlnIleAlaHisAsnAlaGlyLeuGluGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db        460 SerValIleValGluArgLeuLysLysGluGluAlaGlyPheGlyPheAsnAlaAlaThr 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db        480 GlyGluTrpValAsnMetValGluAlaGlyIleValAspProThrLysValThrArgSer 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db        500 AlaLeuGlnHisAlaAlaSerValSerAlaMetPheLeuThrThrGluAlaValIleAla 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGT---------ATGGGT 1617
Db        520 AspLysProGluGluAsnGluGly-----------GlyGlyGlyMetProAspMetGly 535

Qy       1618 GGTATGGGTGGTATGGACGGTATG 1641
Db        536 GlyMetGlyGlyMetGlyGlyMet 543
```

Search completed: January 28, 2004, 13:42:35
Job time : 87.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:23:01 ; Search time 19.5 Seconds
(without alignments)
7147.280 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125847_11658/app_query.fasta_1.1799
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574_@CGN_1_1_27_@runat_28012004_125847_11658 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1844	62.1	548	2	US-08-467-822-31	Sequence 31, Appl
2	1844	62.1	548	3	US-08-432-697-31	Sequence 31, Appl
3	1844	62.1	548	3	US-08-466-248-31	Sequence 31, Appl
4	1820.5	61.3	551	4	US-09-252-991A-24209	Sequence 24209, A
5	1768.5	59.6	550	4	US-09-328-352-5508	Sequence 5508, Ap
6	1741	58.7	548	2	US-08-467-822-32	Sequence 32, Appl
7	1741	58.7	548	3	US-09-472-971-3	Sequence 3, Appli
8	1741	58.7	548	3	US-08-432-697-32	Sequence 32, Appl
9	1741	58.7	548	3	US-08-466-248-32	Sequence 32, Appl
10	1723.5	58.1	546	3	US-08-470-260-6	Sequence 6, Appli
11	1723.5	58.1	546	3	US-08-471-491-6	Sequence 6, Appli
12	1723.5	58.1	546	3	US-08-466-662-6	Sequence 6, Appli
13	1723.5	58.1	546	4	US-08-256-847C-1	Sequence 1, Appli
14	1723.5	58.1	546	4	US-08-256-847C-7	Sequence 7, Appli
15	1721	58.0	544	2	US-08-467-822-33	Sequence 33, Appl
16	1721	58.0	544	3	US-08-432-697-33	Sequence 33, Appl
17	1721	58.0	544	3	US-08-466-248-33	Sequence 33, Appl
18	1716.5	57.8	547	4	US-08-461-722-2	Sequence 2, Appli
19	1716.5	57.8	547	4	US-08-336-251-2	Sequence 2, Appli
20	1716.5	57.8	547	4	US-09-468-041-2	Sequence 2, Appli
21	1716.5	57.8	547	5	PCT-US94-06362-2	Sequence 2, Appli
22	1708	57.5	544	4	US-09-198-452A-153	Sequence 153, App
23	1702	57.3	545	2	US-08-467-822-30	Sequence 30, Appl
24	1702	57.3	545	3	US-08-432-697-30	Sequence 30, Appl
25	1702	57.3	545	3	US-08-466-248-30	Sequence 30, Appl
26	1683.5	56.7	641	4	US-09-613-303-51	Sequence 51, Appl
27	1597	53.8	541	2	US-08-997-080-160	Sequence 160, App
28	1597	53.8	541	2	US-08-997-362-160	Sequence 160, App
29	1597	53.8	541	3	US-09-095-855-160	Sequence 160, App
30	1597	53.8	541	4	US-09-324-542-160	Sequence 160, App
31	1597	53.8	541	4	US-09-205-426-160	Sequence 160, App
32	1595.5	53.8	540	4	US-09-613-303-4	Sequence 4, Appli
33	1595.5	53.8	639	4	US-09-613-303-17	Sequence 17, Appl
34	1595.5	53.8	648	4	US-09-613-303-29	Sequence 29, Appl
35	1595.5	53.8	948	4	US-09-613-303-21	Sequence 21, Appl
36	1591.5	53.6	540	4	US-08-461-722-4	Sequence 4, Appli
37	1591.5	53.6	540	4	US-08-336-251-4	Sequence 4, Appli
38	1591.5	53.6	540	4	US-09-468-041-4	Sequence 4, Appli
39	1591.5	53.6	540	5	PCT-US94-06362-4	Sequence 4, Appli
40	1582.5	53.3	540	4	US-09-134-001C-4705	Sequence 4705, Ap
41	1577.5	53.2	541	2	US-08-467-822-34	Sequence 34, Appl
42	1577.5	53.2	541	2	US-08-447-154-19	Sequence 19, Appl
43	1577.5	53.2	541	3	US-08-432-697-34	Sequence 34, Appl
44	1577.5	53.2	541	3	US-08-466-248-34	Sequence 34, Appl
45	1575	53.1	540	2	US-08-368-834-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-467-822-31
; Sequence 31, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-822-31

Alignment Scores:
Pred. No.:              5.71e-170       Length:          548
Score:                  1844.00         Matches:         366
Percent Similarity:     82.02%          Conservative:    81
Best Local Similarity:  67.16%          Mismatches:      96
Query Match:            62.13%          Indels:          2
DB:                     2               Gaps:            1

US-09-077-574A-1 (1-1647) x US-08-467-822-31 (1-548)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||:::   |||         ||||||   ::::::      ||||||:::
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::||||||:::||||||:::||||||:::||||||||||||||||:::||||||
Db         22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   :::||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              |||   :::|||   ||||||||||||||||||||||||||||||||   ||||||||
Db         62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||||:::::|||   |||
Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||   |||   ||||||||||||   |||||||||   :::||||||||||||||||||
Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||:::|||||||||||||:::|||       ::::::||||||   ::::||   |||
Db        122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              ||||||||||||   ||||||||||||   ||||||||||||||||:::       |||
Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              ||||||      |||   ||||||||||||||:::||||||||||||||||:::||||||
Db        182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||:::   |||   :::   |||   |||||||||:::::|||:::||||||
Db        202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::||||||:::::|||:::::::::||||||   :::||||||   ||||||||
Db        222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
              ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        242 SerGlyArgProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              |||||||||   :::||||||   ::::::|||   ||||||||||||||||||||||
Db        262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::||||||||||||||||||:::||||||||||||||||||   |||:::   |||
Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||       ||||||       :::|||       ||||||:::|||||||||
Db        302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||:::  ||||||||||||||||||:::||||||   ||||||:::   :::|||
Db        322 IleValValThrLysGluAsnThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              ||||||:::   ||||||||||||||:::|||||||||:::|||||||||||||||||
Db        342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||:::||||||||   ||||||||||||||||   ||||||
Db        362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||||   ||||||||||||||||||   ||||||||||||||||:::|||
Db        382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              |||   ||||||   :::|||       :::||||||:::       |||:::||||||
Db        422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::||||||:::::|||   |||:::||||||||       ||||||||||||||
Db        442 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::||||||   ||||||   |||   ||||||   :::|||||||||||||||:::
Db        462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              ||||||||   |||::::::::::   |||:::::||||||   ||||||||||||:::
Db        482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||||||||||||||||||:::|||||||||||||   :::|||
Db        502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
              :::         ||||||||:::      :::      |||   ||||||||||||
Db        522 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539

Qy       1627 GGTATGGACGGTATG 1641
              ||||||   ||||||
Db        540 GlyMetGlyGlyMet 544

RESULT 2
US-08-432-697-31
; Sequence 31, Application US/08432697
; Patent No. 6248330
;  GENERAL INFORMATION:
```

; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-432-697-31

Alignment Scores:

Pred. No.:	5.71e-170	Length:	548
Score:	1844.00	Matches:	366
Percent Similarity:	82.02%	Conservative:	81
Best Local Similarity:	67.16%	Mismatches:	96
Query Match:	62.13%	Indels:	2
DB:	3	Gaps:	1

US-09-077-574A-1 (1-1647) x US-08-432-697-31 (1-548)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||:::   |||         ||||||  ::::::      ||||||:::
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::||||||:::||||||:::||||||:::||||||||||||||||:::||||||
Db         22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   :::||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              |||   :::|||   |||||||||||||||||||||||||||||||||  ||||||||
Db         62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||||:::::|||      |||
Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||   |||   ||||||||||||   ||||||||   :::||||||||||||||||||
Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||:::|||||||||||:::|||      ::::::||||||   :::|||   |||
Db        122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              ||||||||||||   ||||||||||||   ||||||||||||||||||:::      |||
Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              ||||||      |||   ||||||||||||||||:::||||||||||||||||:::||||||
Db        182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||:::   |||   :::   |||   ||||||||:::::|||:::||||||
Db        202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
                 :::||||||:::::|||:::::::::||||||   :::||||||   ||||||||
Db        222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        242 SerGlyArgProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||   :::|||||||   ::::::|||   ||||||||||||||||||||||||
Db        262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::||||||||||||||||:::||||||||||||||||||   |||:::   |||
Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||   ||||||      :::|||      ||||||:::||||||||
Db        302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||:::   ||||||||||||||||||:::||||||   ||||||:::   :::|||
Db        322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              ||||||:::   ||||||||||||||:::|||||||||:::|||||||||||||||||
Db        342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||:::|||||||||   ||||||||||||||||||   ||||||
Db        362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||   ||||||||||||||||||   ||||||||||||||||||:::|||
Db        382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||||||   ||||||||   |||   :::|||:::
Db        402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              |||   ||||||   :::|||      :::||||||:::      |||:::||||||
Db        422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::||||||:::::|||   |||:::||||||||      ||||||||||||||
Db        442 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::||||||   ||||||   |||   ||||||   :::|||||||||||||:::
```

```
Db        462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506

Db        482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566

Db        502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626

Db        522 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539

Qy       1627 GGTATGGACGGTATG 1641

Db        540 GlyMetGlyGlyMet 544

RESULT 3
US-08-466-248-31
; Sequence 31, Application US/08466248
; Patent No. 6258359
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,248
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-248-31
```

```
Alignment Scores:
Pred. No.:                5.71e-170     Length:          548
Score:                    1844.00       Matches:         366
Percent Similarity:       82.02%        Conservative:    81
Best Local Similarity:    67.16%        Mismatches:      96
Query Match:              62.13%        Indels:          2
DB:                       3             Gaps:            1

US-09-077-574A-1 (1-1647) x US-08-466-248-31 (1-548)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66

Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126

Db         22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186

Db         42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246

Db         62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306

Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366

Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426

Db        122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486

Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546

Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606

Db        182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666

Db        202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726

Db        222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786

Db        242 SerGlyArgProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846

Db        262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906

Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966

Db        302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321
```

```
Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||:::   |||||||||||||||||:::|||||   ||||||:::   :::|||
Db        322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
                ||||||:::   |||||||||||||||:::||||||||:::|||||||||||||||
Db        342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||:::|||||||||  |||||||||||||||||  |||||
Db        362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||||  ||||||||||||||||  ||||||||||||||||||:::|||
Db        382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||||||||||||||||||||||||||||  |||||||||  |||  :::|||:::
Db        402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                |||   ||||||   :::|||      :::||||||:::      |||:::||||||
Db        422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::||||||::::::|||   |||:::|||||||||      ||||||||||||||
Db        442 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::||||||   ||||||   |||   ||||||   :::|||||||||||||||:::
Db        462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||||||||   |||:::::::::   |||::::::||||||   ||||||||||:::
Db        482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              |||||||||||||||||||||||||||||:::|||||||||||||||   :::|||
Db        502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
              :::         |||||||||:::      :::      |||   ||||||||||||||
Db        522 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539

Qy       1627 GGTATGGACGGTATG 1641
              ||||||   ||||||
Db        540 GlyMetGlyGlyMet 544
```

RESULT 4
US-09-252-991A-24209
; Sequence 24209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24209
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24209

Alignment Scores:
Pred. No.: 1.08e-167 Length: 551

Score:	1820.50	Matches:	366
Percent Similarity:	80.33%	Conservative:	75
Best Local Similarity:	66.67%	Mismatches:	103
Query Match:	61.34%	Indels:	5
DB:	4	Gaps:	2

US-09-077-574A-1 (1-1647) x US-09-252-991A-24209 (1-551)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              |||||:::|||||:::   |||         ||||||:::|||:::      ||||||
Db          5 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 24

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||:::|||||||||  ||||||||||||||||||||||||||||:::
Db         25 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 44

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||:::|||  |||||||||||||||||||||||||||||||||||
Db         45 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 64

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              |||||:::|||||||||||||||||||||||||||:::||||||:::||||||   |||
Db         65 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 84

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              :::|||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 104

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||:::|||  ||||||||||  |||||||||  :::||||||||||||||||||
Db        105 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 124

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||||||||  ||||||:::      :::|||  ::::::  ||||||      |||
Db        125 AspLysAlaThrValAlaIleValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 144

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||  ||||||||||||||||||||||||||||||||||||  :::||||||
Db        145 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 164

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||  ||||||||||||  ||||||||||||||||||||||||||
Db        165 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 184

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||||||||  |||  ||||||||||||||:::|||||||||||||||||||||||||
Db        185 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 204

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||  |||:::  |||  |||||||||:::|||  :::
Db        205 SerProTyrPheValAsnLysProAspThrMetAlaAlaGluLeuAspSerProLeuLeu 224

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::|||||||||:::::::::::::::|||||||||:::||||||   |||
Db        225 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 244

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      ||||||||||||||:::|||||||||||||||||||||||||||||
Db        245 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 264

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||  :::||||||  ::::::|||  ||||||||||||||||||
Db        265 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 284

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||:::|||||||||||||||||||||||||||||
Db        285 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 304

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::  |||:::  ||||||      :::|||      ||||||   |||
Db        305 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 324

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
```

```
Db        325 LysArgValValIleAsnLysGluAsnThrThrIleIleAspGlyAlaGlyValGlnAla 344

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db        345 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 364

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db        365 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 384

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db        385 ValGlyAlaAlaThrGluValGluMetLysGluAsnLysAlaArgValGluAspAlaLeu 404

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db        405 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 424

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db        425 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGluGlnAsnValGlyIle 444

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db        445 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 464

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db        465 GluProSerValValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 484

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db        485 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 504

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db        505 ArgSerAlaLeuGlnAlaAlaAlaSerIleGlyGlyLeuMetIleThrThrGluAlaMet 524

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGT------GGTATG 1614

Db        525 ValAlaGluIleValGlu---------AspLysProAlaMetGlyGlyMetProAspMet 541

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641

Db        542 GlyGlyMetGlyGlyMetGlyGlyMet 550

RESULT 5
US-09-328-352-5508
; Sequence 5508, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5508
;   LENGTH: 550
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5508

Alignment Scores:
Pred. No.:              1.17e-162       Length:         550
Score:                  1768.50         Matches:        348
Percent Similarity:     79.96%          Conservative:   87
Best Local Similarity:  63.97%          Mismatches:     108
Query Match:            59.59%          Indels:         1
DB:                     4               Gaps:           1

US-09-077-574A-1 (1-1647) x US-09-328-352-5508 (1-550)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db          7 MetSerAlaLysAspValLysPheGlyAspSerAlaArgSerLysMetIleAlaGlyVal 26

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Db         27 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 46

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180

Db         47 AspArgSerPheGlyAlaProHisIleThrLysAspGlyValThrValAlaLysGluIle 66

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240

Db         67 SerLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValArgGluValSerSerLys 86

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Db         87 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleLeu 106

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Db        107 AsnGluGlyIleLysSerValThrAlaGlyMetAsnProMetAspLeuLysArgGlyIle 126

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db        127 AspIleAlaValLysThrValValGluAsnIleArgSerIleAlaLysProAlaAspAsp 146

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db        147 PheLysAlaIleGluGlnValGlySerIleSerAlaAsnSerAspThrThrValGlyLys 166

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db        167 LeuIleAlaGlnAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 186

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db        187 SerGlyPheGluAspAlaLeuAspValValGluGlyMetGlnPheAspArgGlyTyrIle 206

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db        207 SerProTyrPheAlaAsnLysGlnAspThrLeuThrAlaGluLeuGluAsnProPheIle 226

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db        227 LeuLeuValAspLysLysIleSerAsnIleArgGluLeuIleSerValLeuGluAlaVal 246

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db        247 AlaLysThrGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 266

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db        267 ThrLeuValValAsnAsnMetArgGlyIleIleLysValCysAlaValLysAlaProGly 286

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db        287 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyAlaThrVal 306

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db        307 IleSerGluGluValGlyMetSerLeuGluGlnAlaThrLeuGlnAspLeuGlyThrAla 326

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db        327 HisLysIleThrValSerLysGluAsnThrValIleValAspGlyAlaGlyAspAlaAla 346

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db        347 AlaIleAlaGluArgValGlnGlnIleArgAlaGlnIleGluGluSerThrSerGluTyr 366
```

```
Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        367 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 386

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        387 IleGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValAspAspAlaLeu 406

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        407 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuVal 426

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        427 ArgAlaValAsnAlaLeuGluGlyLeuLysGlyAlaAsnGluAspGlnThrAlaGlyIle 446

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        447 AsnIleLeuArgArgAlaIleGluAlaProLeuArgGlnIleValAlaAsnAlaGlyAsp 466

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        467 GluProSerValValIleAsnAlaValLysAsnGlyGluGlyAsnPheGlyTyrAsnAla 486

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        487 AlaThrGlyGluTyrGlyAspMetLeuGluMetGlyIleLeuAspProAlaLysValThr 506

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        507 ArgSerAlaLeuGluHisAlaAlaSerValAlaGlyLeuMetLeuThrThrGluCysMet 526

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        527 IleThrAspIleProGluAspLysProAlaAlaPro---AspMetGlyGlyMetGlyGly 545

Qy       1621 ATGGGTGGTATG 1632
Db        546 MetGlyGlyMet 549
```

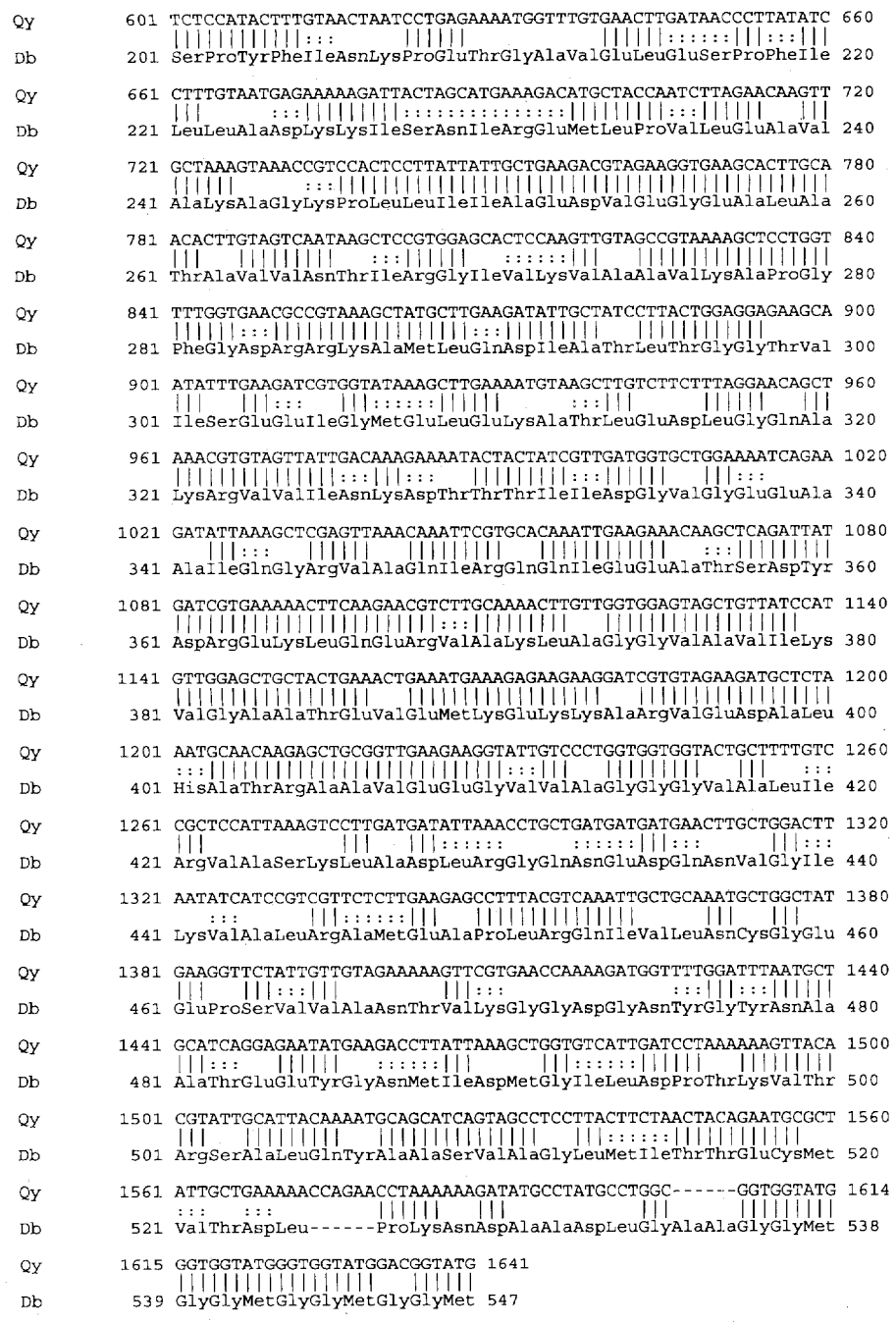
RESULT 6
US-08-467-822-32
; Sequence 32, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-822-32

Alignment Scores:

Pred. No.:	5.37e-160	Length:	548
Score:	1741.00	Matches:	353
Percent Similarity:	77.78%	Conservative:	74
Best Local Similarity:	64.30%	Mismatches:	118
Query Match:	58.66%	Indels:	4
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-467-822-32 (1-548)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200
```



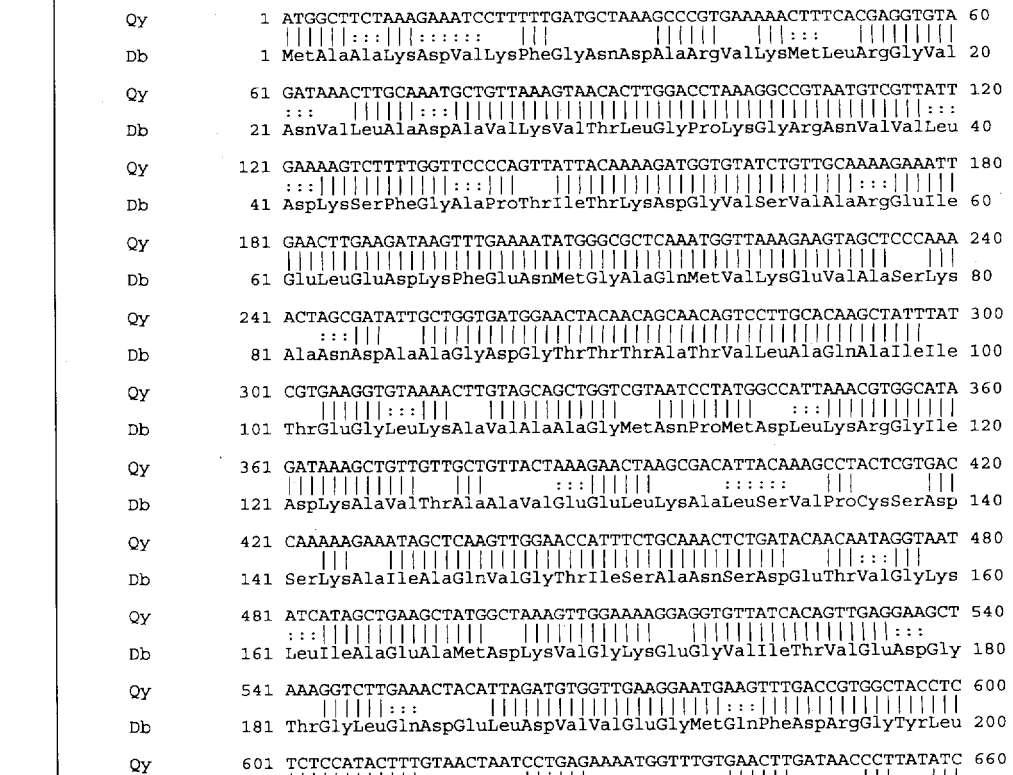
RESULT 7

US-09-472-971-3
; Sequence 3, Application US/09472971
; Patent No. 6197547
; GENERAL INFORMATION:
; APPLICANT: SOGO, Kazuyo
; APPLICANT: YANAGI, Hideki
; APPLICANT: YURA, Takashi
; TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
; FILE REFERENCE: 1422-409P
; CURRENT APPLICATION NUMBER: US/09/472,971
; CURRENT FILING DATE: 1999-12-28
; EARLIER APPLICATION NUMBER: JP10-372965
; EARLIER FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-472-971-3

Alignment Scores:

Pred. No.:	5.37e-160	Length:	548
Score:	1741.00	Matches:	353
Percent Similarity:	77.78%	Conservative:	74
Best Local Similarity:	64.30%	Mismatches:	118
Query Match:	58.66%	Indels:	4
DB:	3	Gaps:	2

US-09-077-574A-1 (1-1647) x US-09-472-971-3 (1-548)



```
Db        201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::|||||||||:::::::::::::::|||||||:::||||||   |||
Db        221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      :::|||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              |||   |||||||||   :::||||||   ::::::|||   ||||||||||||||||
Db        261 ThrAlaValValAsnThrIleArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||:::|||||||||   ||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||::::::||||||      :::|||      ||||||   |||
Db        301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||:::|||:::   |||||||||:::||||||   |||:::
Db        321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||:::   ||||||   |||||||||   ||||||||||||   :::|||||||||
Db        341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||:::|||||||||   |||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||||   ||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||:::|||   |||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||              |||   |||::::::      ::::::|||:::      ||:::
Db        421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 :::      |||::::::|||   ||||||||||||||      |||   |||
Db        441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::|||         |||:::                ::::|||:::||||||
Db        461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::   ||||||   ::::::|||      |||:::::::||||||   |||||||||
Db        481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   |||||||||   |||||||||||||||   |||:::::|||||||||:::
Db        501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614
              :::   :::         ||||||   |||              |||      |||||||||
Db        521 ValThrAspLeu------ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
              ||||||||||||||||   ||||||
Db        539 GlyGlyMetGlyGlyMetGlyGlyMet 547

RESULT 8
US-08-432-697-32
; Sequence 32, Application US/08432697
; Patent No. 6248330
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-432-697-32

Alignment Scores:
Pred. No.:                5.37e-160      Length:          548
Score:                    1741.00        Matches:         353
Percent Similarity:       77.78%         Conservative:    74
Best Local Similarity:    64.30%         Mismatches:      118
Query Match:              58.66%         Indels:          4
DB:                       3              Gaps:            2

US-09-077-574A-1 (1-1647) x US-08-432-697-32 (1-548)

Qy           1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
               ||||||:::|||:::::::   |||          ||||||   |||:::   |||||||||
Db           1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy          61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
               :::   ||||||:::||||||||||||||||||||||||||||||||||||||||:::
Db          21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy         121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
               :::|||||||||||:::|||   |||||||||||||||||||||||||:::||||||
Db          41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy         181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db          61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy         241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
                  :::|||   ||||||||||||||||||||||||||||||||||||||||||||||
Db          81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100
```

```
Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       261 ThrAlaValValAsnThrIleArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy      1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy      1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy      1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614
Db       521 ValThrAspLeu------ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db       539 GlyGlyMetGlyGlyMetGlyGlyMet 547
```

```
RESULT 9
US-08-466-248-32
; Sequence 32, Application US/08466248
; Patent No. 6258359
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,248
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-08-466-248-32

Alignment Scores:

Pred. No.:	5.37e-160	Length:	548
Score:	1741.00	Matches:	353
Percent Similarity:	77.78%	Conservative:	74
Best Local Similarity:	64.30%	Mismatches:	118
Query Match:	58.66%	Indels:	4
DB:	3	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-466-248-32 (1-548)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::  |||       ||||||  |||:::  |||||||||
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::  ||||||:::||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||||||||:::|||  ||||||||||||||||||||||||||:::||||||
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              :::|||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||:::|||  ||||||||||  ||||||||  :::||||||||||||
Db        101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||  |||     :::|||||||     ::::::  |||       |||
Db        121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||  ||||||||||||||||||||||||||||||||||||  |||:::|||
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::|||||||||||||||  ||||||||||  ||||||||||||||||:::
Db        161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||:::  ||||||||||||||||||||||:::||||||||||||||||||
Db        181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::      ||||||              ||||||::::::|||:::|||
Db        201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||       :::|||||||||:::::::::::::::|||||||||:::||||||   |||
Db        221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      :::||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              |||   |||||||||   :::||||||   ::::::|||   ||||||||||||||||||
Db        261 ThrAlaValValAsnThrIleArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||:::|||||||||   ||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||::::::||||||      :::|||      ||||||   |||
Db        301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||:::|||:::   ||||||||||:::||||||   |||:::
Db        321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||:::   ||||||   |||||||||   ||||||||||||   :::||||||||||
Db        341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||:::|||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||  ||||||||||||||||||  ||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||:::|||  ||||||||  |||  :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||              |||  |||::::::      ::::::|||:::      |||:::
Db        421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 :::      |||::::::|||  ||||||||||||||      |||  |||
Db        441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||  |||:::|||          |||:::                 :::|||:::||||||
Db        461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::  ||||||  ::::::|||      |||::::::||||||  |||||||||
Db        481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||  |||||||||  ||||||||||||||  |||::::::||||||||||||
Db        501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614
              :::  :::          ||||||  |||          |||      |||||||||
Db        521 ValThrAspLeu------ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
              ||||||||||||||||||  ||||||
Db        539 GlyGlyMetGlyGlyMetGlyGlyMet 547
```

RESULT 10
US-08-470-260-6
; Sequence 6, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,260
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/256,848
;      FILING DATE:  21-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0316.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2708
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  546 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-470-260-6

Alignment Scores:
Pred. No.:                  2.65e-158      Length:          546
Score:                      1723.50        Matches:         350
Percent Similarity:         77.47%         Conservative:    73
Best Local Similarity:      64.10%         Mismatches:      120
Query Match:                58.07%         Indels:          3
DB:                         3              Gaps:            2

US-09-077-574A-1 (1-1647) x US-08-470-260-6 (1-546)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||||||||   |||       ||||||     |||     ||||||   :::
Db          2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||   :::|||||||||||||||:::||||||:::|||||||||||||:::|||:::|||
Db         22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   ||||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                          ||||||||||||||:::|||||||||||||||   ||||||:::
Db         62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||||||||   :::|||:::::::|||
Db         82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||::::::   :::   ||||||   ||||||:::   :::||||||||||:::||||||
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              |||     |||:::     ||||||        :::|||                ::::::
Db        122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||   ||||||   ||||||||||||||||||||||||     ||||||   :::|||
Db        142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||   ||||||||||||   ||||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||     ||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||   |||||||||      :::|||||||||   |||||||||
Db        202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
                :::|||||||||:::|||||||||||:::|||||:::|||||:::      |||
Db        222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                 :::|||||||||||||||||||||:::|||||||||||||||   ||||||
Db        242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||||||||||||   |||   :::   ||||||||||||||||||||||||
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||   ||||||:::|||||||||||||||||||||||:::   |||
Db        282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||:::   |||||||||      :::      ||||||   |||   |||
Db        302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::||||||||||||:::||||||||||||||||||||||   |||   |||:::|||:::
Db        322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||   ||||||   ||||||:::   ||||||      |||:::|||||||||||||:::
Db        342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||||||||||||||   ||||||||||||||||||   ||||||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||:::|||   ||||||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||:::||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                :::   :::                  ||||||||      |||      |||
Db        422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              |||   |||:::::::::   ||||||   |||||||||   ||||||||||:::|||
Db        440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
                ::::||||||   :::|||   :::   :::      ||||||||||||||:::::
Db        460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||:::|||   |||:::   |||   |||:::|||||||   ||||||   ||||||
Db        480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||   ||||||:::|||||||||||||||||||||||||   :::
Db        500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy       1567 GAAAAACCAGAACCTAAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
              |||      |||   |||                ||||||      ||||||||||||||||||
Db        520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy       1624 GGTGGTATGGACGGTATG 1641
              |||||||||   ||||||
```

```
Db        540 GlyGlyMetGlyGlyMet 545

RESULT 11
US-08-471-491-6
; Sequence 6, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0044
;  CURRENT APPLICATION NUMBER: US/08/471,491B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-08-471-491-6

Alignment Scores:
Pred. No.:                2.65e-158      Length:          546
Score:                    1723.50        Matches:         350
Percent Similarity:       77.47%         Conservative:    73
Best Local Similarity:    64.10%         Mismatches:      120
Query Match:              58.07%         Indels:          3
DB:                       3              Gaps:            2

US-09-077-574A-1 (1-1647) x US-08-471-491-6 (1-546)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||||||||   |||         ||||||      |||      ||||||   :::
Db          2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||   :::|||||||||||||:::||||||:::|||||||||||:::|||:::|||
Db         22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||  .  ||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                          ||||||||||||||:::|||||||||||||||   ||||||:::
Db         62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   |||||||||||||||||||||||||||||||||||||   :::|||::::::|||
Db         82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||::::::   :::   ||||||   ||||||:::   :::|||||||||:::||||||
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              |||      |||:::      ||||||         :::|||              ::::::
Db        122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||   ||||||   ||||||||||||||||||||      ||||||   :::|||
Db        142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||   ||||||||||||   ||||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||      ||||||||||||||||||:::||||||||||||||||||||||||||

Db        182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||   |||||||||      :::|||||||||   |||||||||
Db        202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::|||||||||:::|||||||||||:::||||||:::|||||:::       |||
Db        222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                  :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||
Db        242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||||||||   |||   :::   ||||||||||||||||||||||||
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||   ||||||:::|||||||||||||||||||||:::   |||
Db        282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||:::   ||||||||      :::      ||||||   |||   |||
Db        302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||||||||||:::|||||||||||||||||||||   |||   |||:::|||:::
Db        322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||   ||||||   ||||||:::   ||||||      |||:::||||||||||||:::
Db        342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              |||||:::|||   ||||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||:::|||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              :::   :::                  |||||||||      |||      |||
Db        422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              |||   |||:::::::::   ||||||   |||||||||   ||||||||||||:::|||
Db        440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              :::||||||   :::|||   :::   :::      ||||||||||||||:::::::
Db        460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||:::|||   |||:::   |||   |||:::|||||||||   ||||||   ||||||
Db        480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||   ||||||:::|||||||||||||||||||||      :::
Db        500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
              |||      |||   |||                ||||||      ||||||||||||||||
Db        520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy       1624 GGTGGTATGGACGGTATG 1641
              |||||||||   ||||||
Db        540 GlyGlyMetGlyGlyMet 545
```

RESULT 12
US-08-466-662-6
; Sequence 6, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-6

Alignment Scores:

Pred. No.:	2.65e-158	Length:	546
Score:	1723.50	Matches:	350
Percent Similarity:	77.47%	Conservative:	73
Best Local Similarity:	64.10%	Mismatches:	120
Query Match:	58.07%	Indels:	3
DB:	3	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-466-662-6 (1-546)

```
Qy        7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
            :::|||||||||  |||      ||||||      |||     ||||||   :::
Db        2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy       67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
            |||   :::|||||||||||||||:::||||||:::|||||||||||||:::|||:::|||
Db       22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy      127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
            |||:::|||:::|||  ||||||||||||||||||||||||||||||||||||||||||||
Db       42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy      187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                        ||||||||||||||:::|||||||||||||||   ||||||:::
Db       62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy      247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
            |||   ||||||||||||||||||||||||||||||||   :::|||:::::|||
Db       82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy      307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
            |||::::::   :::   ||||||  ||||||:::   :::|||||||||:::||||||
Db      102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy      367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
            |||     |||:::     ||||||        :::|||        :::::
Db      122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy      427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
            ||||||  ||||||  ||||||||||||||||||||      ||||||  :::|||
Db      142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy      487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
            |||:::||||||  ||||||||||||  ||||||||||||||||||||||||||||||||
Db      162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy      547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
            :::|||     ||||||||||||||||||||:::||||||||||||||||||||||||||
Db      182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy      607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
            ||||||||||||||  ||||||||      :::|||||||||  ||||||||||
Db      202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy      667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
               :::|||||||||:::||||||||||||:::||||||:::||||||:::      |||
Db      222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuGluLysThrMetLys 241

Qy      727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                  :::||||||||||||||||||||||||||:::||||||||||||||||  ||||||
Db      242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy      787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
            ||||||||||||||||||||  |||  :::  ||||||||||||||||||||||||||||
Db      262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy      847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
            :::|||||||||  ||||||:::||||||||||||||||||||||:::  |||
Db      282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy      907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
            |||:::   |||:::  |||||||||      :::     ||||||  |||  |||
Db      302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy      967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
            :::||||||||||||:::||||||||||||||||||||||  |||  |||:::|||:::
Db      322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy     1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
            |||  ||||||  ||||||:::  ||||||     |||:::||||||||||||:::
Db      342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy     1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
            ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db      362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy     1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
            ||||||:::|||  ||||||||||||||||||||||||||:::|||||||||:::|||
Db      382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy     1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
            |||:::|||||||||||||||||||||||  |||||||||  |||  :::|||:::
Db      402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy     1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
               :::  :::                  |||||||||  |||  |||
Db      422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy     1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
            |||   |||:::::::::  ||||||  |||||||||  ||||||||||||:::|||
Db      440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy     1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
               :::||||||  :::|||  :::  :::  ||||||||||||||:::::
Db      460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy     1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
            |||:::|||  |||:::  |||  |||:::||||||||  ||||||  ||||||
Db      480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy     1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
            ||||||||||||||  ||||||:::||||||||||||||||||||||  :::
Db      500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy     1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
            |||     |||  |||              ||||||  ||||||||||||||||||
Db      520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy     1624 GGTGGTATGGACGGTATG 1641
            |||||||||  ||||||
Db      540 GlyGlyMetGlyGlyMet 545
```

RESULT 13
US-08-256-847C-1
; Sequence 1, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
; APPLICANT: Rappuoli, Rino
; APPLICANT: Costantino, Paolo
; APPLICANT: No. 6403099elli, Francesco
; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
; FILE REFERENCE: CHIR-0042
; CURRENT APPLICATION NUMBER: US/08/256,847C
; CURRENT FILING DATE: 1994-11-01
; PRIOR APPLICATION NUMBER: PCT/EP93/00516
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: FI92A000058
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 546
; TYPE: PRT
; ORGANISM: H. pylori
US-08-256-847C-1

Alignment Scores:

Pred. No.:	2.65e-158	Length:	546
Score:	1723.50	Matches:	350
Percent Similarity:	77.47%	Conservative:	73
Best Local Similarity:	64.10%	Mismatches:	120
Query Match:	58.07%	Indels:	3
DB:	4	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-256-847C-1 (1-546)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||||||||||   |||         ||||||      |||     ||||||  :::
Db          2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||   :::|||||||||||||||:::||||||:::|||||||||||||:::|||:::|||
Db         22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   ||||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                          ||||||||||||||:::|||||||||||||||   ||||||:::
Db         62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||||   :::|||:::::|||
Db         82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||::::::   :::   ||||||   ||||||:::   :::||||||||||:::||||||
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              |||      |||:::      ||||||         :::|||          ::::::
Db        122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||   ||||||   ||||||||||||||||||||      ||||||   :::|||
Db        142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||   ||||||||||||   ||||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||     ||||||||||||||||||:::||||||||||||||||||||||||||||
Db        182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||   |||||||||      :::|||||||||   |||||||||
Db        202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
                 :::|||||||||:::|||||||||||||:::|||||:::|||||:::      |||
Db        222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                    :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||
Db        242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||||||||||   |||   :::   ||||||||||||||||||||||||
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
                 :::|||||||||   |||||:::|||||||||||||||||||||||:::   |||
Db        282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||:::   |||||||||      :::      ||||||   |||   |||
Db        302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||||||||||:::|||||||||||||||||||   |||   |||:::|||:::
Db        322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||   ||||||   ||||||:::   ||||||      |||:::|||||||||||:::
Db        342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              |||||||||||||||||||||||||||||||   ||||||||||||||||   ||||||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||:::|||   |||||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||:::|||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              :::   :::                        |||||||||   |||   |||
Db        422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              |||   |||::::::::   ||||||   |||||||||   ||||||||||||:::|||
Db        440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
                 :::||||||   :::|||   :::   :::      ||||||||||||||:::::
Db        460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||:::|||   |||:::   |||   |||:::|||||||||   ||||||   ||||||
Db        480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||   ||||||:::||||||||||||||||||||||      :::
Db        500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
              |||      |||   |||              ||||||      ||||||||||||||||||
Db        520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy       1624 GGTGGTATGGACGGTATG 1641
              |||||||||   ||||||
Db        540 GlyGlyMetGlyGlyMet 545
```

RESULT 14
US-08-256-847C-7
; Sequence 7, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
; APPLICANT: Rappuoli, Rino
; APPLICANT: Costantino, Paolo
; APPLICANT: No. 6403099elli, Francesco
; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
; FILE REFERENCE: CHIR-0042
; CURRENT APPLICATION NUMBER: US/08/256,847C
; CURRENT FILING DATE: 1994-11-01
; PRIOR APPLICATION NUMBER: PCT/EP93/00516
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: FI92A000058
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 546
; TYPE: PRT
; ORGANISM: H. pylori
US-08-256-847C-7

Alignment Scores:

Pred. No.:	2.65e-158	Length:	546
Score:	1723.50	Matches:	350
Percent Similarity:	77.47%	Conservative:	73
Best Local Similarity:	64.10%	Mismatches:	120
Query Match:	58.07%	Indels:	3
DB:	4	Gaps:	2

US-09-077-574A-1 (1-1647) x US-08-256-847C-7 (1-546)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||||||||  |||       ||||||     |||     ||||||   :::
Db          2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||   :::||||||||||||||:::|||||:::|||||||||||||:::|||:::|||
Db         22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||  ||||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                              ||||||||||||||:::|||||||||||||||   ||||||:::
Db         62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   |||||||||||||||||||||||||||||||||||   :::|||:::::|||
Db         82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||::::::   :::   ||||||   ||||||:::   :::||||||||||:::||||||
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              |||     |||:::     ||||||      :::|||                 ::::::
Db        122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||  ||||||  ||||||||||||||||||||      ||||||  :::|||
Db        142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||  ||||||||||||  ||||||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||     ||||||||||||||||||||:::|||||||||||||||||||||||||
Db        182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||||  |||||||||      :::|||||||||  |||||||||
Db        202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
                 :::|||||||||:::||||||||||||:::|||||:::|||||:::     |||
Db        222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                    :::||||||||||||||||||||||||||:::|||||||||||||||  ||||||
Db        242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||||||||||  |||  :::  ||||||||||||||||||||||||||
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||  ||||||:::|||||||||||||||||||||||:::  |||
Db        282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||:::   |||||||||      :::     ||||||  |||  |||
Db        302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||||||||||:::|||||||||||||||||||||||  |||  |||:::|||:::
Db        322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGTCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||   ||||||  ||||||:::  ||||||     |||:::|||||||||||||:::
Db        342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||:::|||  |||||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||:::|||||||||||||||||||||||  |||||||||  |||  :::|||:::
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              :::  :::                  |||||||||  |||  |||
Db        422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              |||   |||:::::::::  ||||||  |||||||||  ||||||||||||:::|||
Db        440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              :::||||||  :::|||  :::  :::  ||||||||||||||:::::
Db        460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||:::|||  |||:::  |||  |||:::|||||||||  ||||||  ||||||
Db        480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||  ||||||:::||||||||||||||||||||||  :::
Db        500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
              |||   |||  |||  ||||||  ||||||||||||||||||||||||||||||||||
Db        520 GluIleLysGluGluLysAlaThrProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy       1624 GGTGGTATGGACGGTATG 1641
              |||||||||  ||||||
Db        540 GlyGlyMetGlyGlyMet 545
```

```
RESULT 15
US-08-467-822-33
; Sequence 33, Application US/08467822
; Patent No. 5843460
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,822
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  544 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-822-33

Alignment Scores:
Pred. No.:                4.62e-158      Length:          544
Score:                    1721.00        Matches:         351
Percent Similarity:       78.23%         Conservative:    73
Best Local Similarity:    64.76%         Mismatches:      114
Query Match:              57.99%         Indels:          4
DB:                       2              Gaps:            2

US-09-077-574A-1 (1-1647) x US-08-467-822-33 (1-544)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||   |||   ::::::      ||||||:::|||:::   :::||||||
Db          1 MetAlaAlaLysAsnIleLysTyrAsnGluAspAlaArgLysLysIleHisLysGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                     ||||||   ||||||||||||||||||||||||||||||:::||||||||
Db         21 LysThrLeuAlaGluAlaValLysValThrLeuGlyProLysGlyArgHisValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||||||   :::|||||||||||||||:::|||||||||||||||
Db         41 AspLysSerPheGlySerProGlnValThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||   |||||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuGluAspLysHisGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||   |||||||||||||||||||||||||||||||||:::||||||||||
Db         81 ThrAlaAspLysAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGluAlaIleTyr 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||::::::   |||   ||||||   |||||||||       ||||||||||||
Db        101 SerGluGlyLeuArgAsnValThrAlaGlyAlaAsnProMetLeuAspLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||   |||:::   |||:::||||||   :::
Db        121 AspLysAlaValLysValValValAspGluIleLysLysIleSerLysProValGlnHis 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||||||||||||||||   |||||||||||||||:::|||       |||||||||
Db        141 HisLysGluIleAlaGlnValAlaThrIleSerAlaAsnAsnAspAlaGluIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::||||||||||||||   ||||||||||||   |||   ||||||||||||||||||
Db        161 LeuIleAlaGluAlaMetGluLysValGlyLysAsnGlySerIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||   ||||||   ||||||||||||||||||||   |||:::||||||||||||
Db        181 LysGlyPheGluThrValLeuAspValValGluGlyMetAsnPheAsnArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              |||   ||||||   ||||||||||||          |||   |||:::         :::
Db        201 SerSerTyrPheSerThrAsnProGluThrGlnGluCysValLeuGluGluAlaLeuVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||       :::||||||||:::   :::||||||   ||||||:::|||:::||||||
Db        221 LeuIleTyrAspLysLysIleSerGlyIleLysAspPheLeuProValLeuGlnGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              |||:::      ||||||||||||||||||||||||||:::|||||||||||||||||
Db        241 AlaGluSerGlyArgProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||:::||||||          :::|||   ||||||||||||||||||
Db        261 ThrLeuValValAsnArgLeuArgAlaGlyPheArgValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnLeu 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::||||||||||||   :::|||:::   ||||||   |||
Db        301 IleSerGluGluLeuGlyMetLysLeuGluAsnThrThrLeuAlaMetLeuGlyLysAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||:::::   ||||||:::||||||||||||:::|||   |||       |||
Db        321 LysLysValIleValSerLysGluAspThrThrIleValGluGlyLeuGlySerLysGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||::::::|||   :::   |||:::   |||||||||:::::::::|||||||||
Db        341 AspIleGluSerArgCysGluSerIleLysLysGlnIleGluAspSerThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||:::|||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              |||||||||||||||||   |||||||||||||||||||||||||:::|||||||
Db        381 ValGlyAlaAlaThrGluIleGluMetLysGluLysLysAspArgValAspAspAlaGln 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
```

```
Db        401 HisAlaThrLeuAlaAlaValGluGluGlyIleLeuProGlyGlyGlyThrAlaLeuVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCT------GCTGATGATGATGAACTTGCT 1314
Db        421 ArgCysIleProThrLeuGluAlaPheIleProIleLeuThrAsnGluAspGluGlnIle 440

Qy       1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCT 1374
Db        441 GlyAlaArgIleValLeuLysAlaLeuSerAlaProLeuLysGlnIleAlaAlaAsnAla 460

Qy       1375 GGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTT 1434
Db        461 GlyLysGluGlyAlaIleIleCysGlnGlnValLeuSerArgSerSerSerGluGlyTyr 480

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db        481 AspAlaLeuArgAspAlaTyrThrAspMetIleGluAlaGlyIleLeuAspProThrLys 500

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db        501 ValThrArgCysAlaLeuGluSerAlaAlaSerValAlaGlyLeuLeuLeuThrThrGlu 520

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATG------CCTATGCCTGGCGGT 1608
Db        521 AlaLeuIleAlaAspIleProGluGluLysSerSerSerAlaProAlaMetProGlyAla 540

Qy       1609 GGTATG 1614
Db        541 GlyMet 542
```

Search completed: January 28, 2004, 13:30:55
Job time : 46.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:17:56 ; Search time 52 Seconds
(without alignments)
10054.718 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125845_11596/app_query.fasta_1.1799
-DB=A_Geneseq_19Jun03 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574_@CGN_1_1_81_@runat_28012004_125845_11596 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2708	91.2	548	18	AAW16678	Lawsonia intracell
2	1864.5	62.8	545	20	AAY23930	Consensus mino aci
3	1848	62.3	549	20	AAY23909	Amino acid sequenc
4	1825.5	61.5	547	22	AAB69060	Pseudomonas aerugi
5	1820.5	61.3	547	16	AAR67381	L. pneumophila Htp
6	1819	61.3	574	22	AAB69061	pTrcHisB expressio
7	1814.5	61.1	545	20	AAY23915	Amino acid sequenc
8	1812.5	61.1	544	21	AAY75747	Neisseria meningit
9	1805.5	60.8	544	24	ABP77245	N. gonorrhoeae ami
10	1769.5	59.6	544	23	AAE22070	Francisella tulare
11	1750.5	59.0	544	21	AAY75745	Neisseria gonorrhe
12	1749	58.9	545	23	ABB99014	Detrimental organi
13	1746	58.8	545	21	AAB19080	Amino acid sequenc
14	1746	58.8	548	19	AAW65070	E. coli GroEL prot
15	1746	58.8	548	20	AAW95001	WO9902989 Seq ID 1
16	1746	58.8	548	22	AAB50536	Escherichia coli G
17	1746	58.8	548	23	ABB99010	Detrimental organi
18	1744	58.8	545	23	ABB99013	Detrimental organi
19	1743.5	58.7	544	20	AAY23905	Amino acid sequenc
20	1742	58.7	545	23	ABB99009	Detrimental organi
21	1742	58.7	545	23	ABB99012	Detrimental organi
22	1742	58.7	545	23	ABB99016	Detrimental organi
23	1741	58.7	548	19	AAW61378	GroEL N-terminal r
24	1741	58.7	548	21	AAY99621	Escherichia coli g
25	1739	58.6	545	23	ABB99015	Detrimental organi
26	1738	58.6	545	23	ABB99011	Detrimental organi
27	1736.5	58.5	548	20	AAY23908	Amino acid sequenc
28	1735	58.5	548	16	AAR67382	E. coli GroEL gene
29	1734	58.4	545	23	ABB99017	Detrimental organi
30	1733	58.4	545	20	AAY23901	Streptococcus pneu
31	1732	58.4	542	23	ABB49241	Listeria monocytog
32	1729.5	58.3	544	20	AAY23903	Streptococcus pyog
33	1723.5	58.1	546	14	AAR41661	H. pylori hsp60.
34	1723.5	58.1	546	14	AAR41200	Hsp. Helicobacter
35	1723	58.1	544	12	AAR13337	HypB protein. Chl
36	1720.5	58.0	546	19	AAW98776	H. pylori GHPO 118
37	1720	58.0	548	23	ABB99019	Detrimental organi
38	1720	58.0	548	23	ABB99021	Detrimental organi
39	1719	57.9	548	23	ABB99018	Detrimental organi
40	1718	57.9	544	16	AAR67383	C. psittaci HypB g
41	1717	57.9	548	23	ABB99020	Detrimental organi
42	1716.5	57.8	547	23	AAU76192	Human P1 protein (
43	1716.5	57.8	547	24	ABG74586	E. coli groEL prot
44	1710	57.6	544	22	AAE11757	Chlamydophila pneu
45	1710	57.6	544	23	ABB94272	Chlamydia pneumoni

ALIGNMENTS

RESULT 1
AAW16678
ID AAW16678 standard; Protein; 548 AA.
XX
AC AAW16678;
XX
DT 20-AUG-1997 (first entry)
XX
DE Lawsonia intracellularis GroEL.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine;
KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.
XX
OS Lawsonia intracellularis.
XX
PN WO9720050-A1.
XX
PD 05-JUN-1997.

XX
PF 29-NOV-1996; 96WO-AU00767.
XX
PR 30-NOV-1995; 95AU-0006911.
PR 30-NOV-1995; 95AU-0006910.
XX
PA (DARA-) DARATECH PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Hasse D, Panaccio M;
XX
DR WPI; 1997-310605/28.
DR N-PSDB; AAT69201.
XX
PT Vaccine for treating or preventing Lawsonia intracellularis
PT infection - especially in pigs, containing non-pathogenic form of
PT bacterium or its components
XX
PS Claim 10; Page 38-42; 94pp; English.
XX
CC The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis
CC can be used in vaccines to protect birds and animals against
CC intestinal diseases, esp. to protect pigs against porcine
CC proliferative enteropathy. It is the expression product of a DNA
CC molecule (AAT69201) obtd. by screening an L. intracellularis library
CC with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)
CC and other L. intracellularis polypeptides (AAW16680-85) can be
CC administered as recombinant polypeptides or expressed as
CC recombinant vaccines utilising bacterial, fungal or viral vectors.
CC Antibodies raised against the polypeptides may be useful in
CC immunotherapy, diagnosis of infection and detection.
XX
SQ Sequence 548 AA;

Alignment Scores:

Pred. No.:	6.89e-236	Length:	548
Score:	2708.00	Matches:	548
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	91.24%	Indels:	0
DB:	18	Gaps:	0

US-09-077-574A-1 (1-1647) x AAW16678 (1-548)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaSerLysGluIleLeuPheAspAlaLysAlaArgGluLysLeuSerArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLysLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GluLysSerPheGlySerProValIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaProLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrSerAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleTyr 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgGluGlyValLysLeuValAlaAlaGlyArgAsnProMetAlaIleLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLysAlaValValAlaValThrLysGluLeuSerAspIleThrLysProThrArgAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnLysGluIleAlaGlnValGlyThrIleSerAlaAsnSerAspThrThrIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleIleAlaGluAlaMetAlaLysValGlyLysGlyGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysGlyLeuGluThrThrLeuAspValValGluGlyMetLysPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetValCysGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LeuCysAsnGluLysLysIleThrSerMetLysAspMetLeuProIleLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysValAsnArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyAlaLeuGlnValValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheGlyGluArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluAla 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IlePheGluAspArgGlyIleLysLeuGluAsnValSerLeuSerSerLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LysArgValValIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysSerGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspIleLysAlaArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuValGlyGlyValAlaValIleHis 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluThrGluMetLysGluLysLysAspArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaPheVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ArgSerIleLysValLeuAspAspIleLysProAlaAspAspAspGluLeuAlaGlyLeu 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AsnIleIleArgArgSerLeuGluGluProLeuArgGlnIleAlaAlaAsnAlaGlyTyr 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GluGlySerIleValValGluLysValArgGluProLysAspGlyPheGlyPheAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AlaSerGlyGluTyrGluAspLeuIleLysAlaGlyValIleAspProLysLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        501 ArgIleAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuLeuThrThrGluCysAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620

Db        521 IleAlaGluLysProGluProLysLysAspMetProMetProGlyGlyGlyMetGlyGly 540

Qy       1621 ATGGGTGGTATGGACGGTATGTAC 1644

Db        541 MetGlyGlyMetAspGlyMetTyr 548
```

RESULT 2
AAY23930
ID AAY23930 standard; Protein; 545 AA.
XX
AC AAY23930;
XX
DT 22-SEP-1999 (first entry)
XX
DE Consensus mino acid sequence of a heat shock protein.
XX
KW Heat shock protein; Hsp; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS Synthetic.
XX
PN WO9935270-A1.
XX
PD 15-JUL-1999.
XX
PF 29-DEC-1998; 98WO-CA01203.
XX
PR 31-DEC-1997; 97US-0001737.
XX
PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI Mizzen L, Wisniewski J;
XX
DR WPI; 1999-430397/36.
XX
PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
PS Disclosure; Fig 10A-E; 176pp; English.
XX
CC AAY23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.
XX
SQ Sequence 545 AA;

Alignment Scores:

Pred. No.:	1.19e-159	Length:	545
Score:	1864.50	Matches:	380
Percent Similarity:	83.15%	Conservative:	74
Best Local Similarity:	69.60%	Mismatches:	85
Query Match:	62.82%	Indels:	7
DB:	20	Gaps:	5

US-09-077-574A-1 (1-1647) x AAY23930 (1-545)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66

Db          2 AlaLysAspIleLysPheGlyGluGluAlaArgArgAlaMetLeuArgGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126

Db         22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186

Db         42 SerPheGlyAlaProThrIleThrLysAspGlyValThrValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246

Db         62 GluAspLysPheGluAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306

Db         82 AspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleValLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366

Db        102 GlyLeuLysAsnValAlaAlaGlyAlaAsnProMetAspLeuArgArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426

Db        122 AlaValAspAlaValValGluGluLeuLysAlaIleAlaLysProValGluThrLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486

Db        142 GluIleAlaGlnValAlaThrIleSerAlaAsnGlyAspGluGluIleGlyGluLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546

Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGlyLysThr 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606

Db        182 LeuGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666

Db        202 TyrPheIleThrAspSerGluLysGlnLysAlaGluLeuGluAspProLeuIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726

Db        222 ThrAspLysLysIleSerAsnIleGlnAspLeuLeuProValLeuGluGluValAlaGln 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786

Db        242 AlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846

Db        262 ValValAsnLysLeuArgGlyThrLeuLysValValAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906

Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966

Db        302 GluGluLeuGlyLeuSerLeuGluAspAlaThrLeuGluAspLeuGlyGlnAlaLysLys 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026

Db        322 ValValValThrLysAspAspThrThrIleValAspGlyAlaGlyAspAlaAla---Ile 340

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086

Db        341 AlaGlyArgValAlaGlnIleArgSerGlnIleGluGluSerThrSerAspTyrAspLys 360

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
```

```
Db        361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 380

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||    |||:::|||||:::|||||||||||||||||||||||||||||||
Db        381 AlaAlaThrGluValGluLeuLysGluArgLysAspArgValGluAspAlaLeuAsnAla 400

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||||||||||||    |||   :::|||:::
Db        401 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuLeuArgAla 420

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                    ||||||   :::|||     :::   |||||      |||:::||||||
Db        421 AlaProAlaLeuAspLysLeuLysThrGluAsnGlyAspGluAlaThrGlyValAsnIle 440

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::   |||:::||||||   |||||||||||||||||   |||||||||   ||||||
Db        441 ValLeuArgAlaLeuGluAlaProLeuArgGlnIleAlaGluAsnAlaGlyLeuGluGly 460

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::|||||||||||||||:::      :::   |||   |||:::|||||||||:::
Db        461 SerValValValGluLysValLysAsnSerGluAlaGly---GlyTyrAsnAlaAlaThr 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              ||||||||   |||:::|||   ||||||:::|||||||||   ||||||||||||
Db        480 GlyGluTyrValAspMetIleAlaAlaGlyIleIleAspProThrLysValThrArgSer 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||||||||||||||||||:::||||||||||||      :::
Db        500 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluAlaValValVal 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCT------ATGCCTGGC------GGTGGTATG 1614
              :::|||||||||   |||:::      |||      ||||||||      |||||||||
Db        520 AspLysProGlu---LysGluAlaAlaProAlaGlyMetProGlyMetMetGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATG 1632
              ||||||||||||||||||
Db        539 GlyGlyMetGlyGlyMet 544
```

RESULT 3
AAY23909

```
ID   AAY23909 standard; Protein; 549 AA.
XX
AC   AAY23909;
XX
DT   22-SEP-1999  (first entry)
XX
DE   Amino acid sequence of a heat shock protein.
XX
KW   Heat shock protein; Hsp; immune response; immunological carrier;
KW   cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS   Legionella pneumophila.
XX
PN   WO9935270-A1.
XX
PD   15-JUL-1999.
XX
PF   29-DEC-1998;   98WO-CA01203.
XX
PR   31-DEC-1997;   97US-0001737.
XX
PA   (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI   Mizzen L,  Wisniewski J;
XX
DR   WPI; 1999-430397/36.
XX
PT   New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT   useful in vaccines, as carriers for other immunogens, as anticancer
PT   agents and for diagnosis
XX
PS   Disclosure; Fig 10A-E; 176pp; English.
XX
CC   AAY23905-30 represent heat shock proteins (Hsps). The specification
CC   describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC   fragments, variants and fusion proteins, are used to elicit or enhance
CC   an immune response against Streptococcus, and to elicit a similar
CC   response to a target antigen fused to the protein. Unlike other
CC   immunological carriers, Hsp60 proteins are not immunosuppressive so
CC   provide an increased response to any conjugated or fused antigen. Also,
CC   where used for cancer control, they lack the side effects associated
CC   with endotoxins. They can also be used to detect specific antibodies
CC   and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC   breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC   polynucleotide is used for recombinant production of the protein, as
CC   a source of primers and probes for detecting streptococci in standard
CC   hybridization/amplification assays, and therapeutically in gene
CC   therapy vectors.
XX
SQ   Sequence   549 AA;
```

Alignment Scores:

Pred. No.:	3.7e-158	Length:	549
Score:	1848.00	Matches:	367
Percent Similarity:	82.02%	Conservative:	80
Best Local Similarity:	67.34%	Mismatches:	96
Query Match:	62.26%	Indels:	2
DB:	20	Gaps:	1

US-09-077-574A-1 (1-1647) x AAY23909 (1-549)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||:::   |||        ||||||   ::::::      ||||||:::
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::|||||:::||||||:::||||||:::||||||||||||||||:::||||||
Db         22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   :::|||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              |||   :::|||   |||||||||||||||||||||||||||||   |||||||||
Db         62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   |||||||||||||||||||||||||||||||||:::::|||      |||
Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||   |||   |||||||||||   |||||||||   :::|||||||||||||||||
Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||:::|||||||||||:::|||      ::::::||||||   :::|||   |||
Db        122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              |||||||||||||||||||||||||||||||||||||||||      ||||||   ||||||
Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              ||||||||||||   ||||||||||||   ||||||||||||||||||:::      |||
Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              ||||||      |||   ||||||||||||||:::||||||||||||||:::||||||
Db        182 LeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||:::   |||   :::   |||   |||||||||:::::|||:::|||||||
Db        202 TyrPheIleAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221
```

```
Qy       667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db       222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

Qy       727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db       242 SerGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261

Qy       787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db       262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy       847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db       282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

Qy       907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db       302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

Qy       967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db       322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy      1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db       342 AsnAlaArgIleThrGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 361

Qy      1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db       362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy      1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db       382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy      1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db       402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy      1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db       422 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 441

Qy      1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db       442 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 461

Qy      1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db       462 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 481

Qy      1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db       482 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 501

Qy      1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db       502 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 521

Qy      1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
Db       522 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 539

Qy      1627 GGTATGGACGGTATG 1641
Db       540 GlyMetGlyGlyMet 544
```

RESULT 4
AAB69060
ID AAB69060 standard; Protein; 547 AA.
XX
AC AAB69060;
XX
DT 18-APR-2001 (first entry)
XX
DE Pseudomonas aeruginosa groEL protein sequence.
XX
KW Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW diagnosis; detection; infection; immune response.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200102577-A1.
XX
PD 11-JAN-2001.
XX
PF 03-JUL-2000; 2000WO-GB02554.
XX
PR 01-JUL-1999; 99GB-0015419.
XX
PA (PROV-) PROVALIS UK LTD.
XX
PI Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;
XX
DR WPI; 2001-080988/09.
DR N-PSDB; AAF32452.
XX
PT Antigenic Pseudomonas aeruginosa proteins, useful in the detection
PT and/or diagnosis of P. aeruginosa infections and for producing vaccines
PT against P. aeruginosa -
XX
PS Claim 3; Fig 25; 129pp; English.
XX
CC The present invention describes antigenic Pseudomonas aeruginosa
CC proteins (P1). The P. aeruginosa proteins have antibacterial activity
CC and can be used in vaccines and as antagonists. The proteins or their
CC fragments, or antibodies are useful in the detection and/or diagnosis
CC of P. aeruginosa. They are also useful for producing a vaccine and
CC inducing an immune response against P. aeruginosa infection. An agent
CC capable of antagonising, inhibiting or otherwise interfering with the
CC function or expression of P1 are useful in the manufacture of a
CC medicament for the treatment or prophylaxis of P. aeruginosa infections.
CC The present sequence represents a specifically claimed P. aeruginosa
CC groEL protein sequence from the present invention.
XX
SQ Sequence 547 AA;

Alignment Scores:

Pred. No.:	3.99e-156	Length:	547
Score:	1825.50	Matches:	367
Percent Similarity:	80.51%	Conservative:	75
Best Local Similarity:	66.85%	Mismatches:	102
Query Match:	61.51%	Indels:	5
DB:	22	Gaps:	2

US-09-077-574A-1 (1-1647) x AAB69060 (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
```

```
Db        81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Db       101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db       121 AspLysAlaThrValAlaIleValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db       141 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db       161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db       181 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db       201 SerProTyrPheValAsnLysProAspThrMetAlaAlaGluLeuAspSerProLeuLeu 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db       221 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db       241 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db       261 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db       281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db       301 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db       321 LysArgValValIleAsnLysGluAsnThrThrIleIleAspGlyAlaGlyValGlnAla 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db       341 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db       361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db       381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db       401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db       421 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGluGlnAsnValGlyIle 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db       441 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db       461 GluProSerValValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 480

Qy      1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db       481 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 500

Qy      1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db       501 ArgSerAlaLeuGlnAlaAlaAlaSerIleGlyGlyLeuMetIleThrThrGluAlaMet 520

Qy      1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGT------GGTATG 1614

Db       521 ValAlaGluIleValGlu---------AspLysProAlaMetGlyGlyMetProAspMet 537

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641

Db       538 GlyGlyMetGlyGlyMetGlyGlyMet 546
```

RESULT 5
AAR67381
ID AAR67381 standard; Protein; 547 AA.
XX
AC AAR67381;
XX
DT 25-MAR-2003 (updated)
DT 22-JUN-1995 (first entry)
XX
DE L. pneumophila HtpB gene product.
XX
KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;
KW GroEL-like protein; Helicobacter felis.
XX
OS Legionella pneumophila.
XX
PN WO9426901-A1.
XX
PD 24-NOV-1994.
XX
PF 19-MAY-1994; 94WO-EP01625.
XX
PR 19-MAY-1993; 93EP-0401309.
PR 19-NOV-1993; 93WO-EP03259.
XX
PA (INRM) INST NAT SANTE & RECH MEDICALE.
PA (INSP) INST PASTEUR.
XX
PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;
XX
DR WPI; 1995-006797/01.
XX
PT DNA from Helicobacter pylori and Helicobacter felis - used to
PT develop prods. for detection, treatment and prevention of
PT Helicobacter infection
XX
PS Disclosure; Fig. 7A(i-vii); 168pp; English.
XX
CC The sequence of the Helicobacter pylori heat shock protein A
CC (given in AAR67374) was compared to that of other GroEL-like
CC proteins from Legionella pneumophila (AAR67381), Escherichia coli
CC (AAR67382), Chlamydia psittaci (AAR67383), Mycobacterium leprae
CC (AAR67384) and human mitochondrial protein P1 (AAR67385), and regions
CC of homology were identified.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 547 AA;

Alignment Scores:

Pred. No.:	1.13e-155	Length:	547
Score:	1820.50	Matches:	363
Percent Similarity:	81.65%	Conservative:	82

Best Local Similarity: 66.61% Mismatches: 97
Query Match: 61.34% Indels: 3
DB: 16 Gaps: 2

US-09-077-574A-1 (1-1647) x AAR67381 (1-547)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||||||:::   |||         ||||||   ::::::      ||||||:::
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgLeuGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              ||||||:::||||||:::||||||:::||||||:::|||||||||||||||:::||||||
Db         22 LeuAlaAspAlaValGlnValThrMetGlyProArgGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   :::||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              |||   :::|||   |||||||||||||||||||||||||||||||||   |||||||||
Db         62 GluHisArgPheMetAsnMetGlyAlaGlnMetValLysGluValAlaSerLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   ||||||||||||||||||||||||||||||||||:::::|||     |||
Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleLeuValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||   |||   ||||||||||||   |||||||||   :::|||||||||||||||||||
Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||:::||||||||||||:::|||      :::::||||||   :::|||   |||
Db        122 AlaValLeuAlaValThrLysLysLeuGlnAlaMetSerLysProCysLysAspSerLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||||||||||||||||||||||||||||||||||   ||||||   ||||||
Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluAlaIleGlyAlaIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              ||||||||||||   ||||||||||||   ||||||||||||||||:::      |||
Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              ||||||      |||   ||||||||||||||||:::|||||||||||||||:::||||||
Db        182 LeuGluAsnGluLeuTyrValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||:::   |||   :::   |||   |||||||||:::::|||:::||||||
Db        202 TyrPheIleAsnAsnGlnGlnAsnMetSerCysGluLeuGluHisProPheIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::||||||:::::::|||:::::::::|||||||   :::|||||||   ||||||||
Db        222 ValAspLysLysValSerSerIleArgGluMetLeuSerValLeuGluGlyValAlaLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
              |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        242 SerGlyArgProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              |||||||||   :::||||||   ::::::|||   ||||||||||||||||||||||
Db        262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||||||||||:::|||||||||||||||||||   |||:::   |||
Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrLysGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||      ||||||      :::|||      ||||||:::||||||||
Db        302 GluGluIleGlyLysSerLeuGluGlyAlaThrLeuGluAspLeuGlySerAlaLysArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::::::::      |||   ||||||||||:::|||||||   |||||||:::   :::|||
Db        322 IleIleValThrAspGlu---ThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 340

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              ||||||:::   ||||||||||||||||:::|||||||||:::|||||||||||||||||
Db        341 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrThrSerAspTyrAspArg 360

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              ||||||||||||||||||:::|||||||||   ||||||||||||||||||   ||||||
Db        361 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 380

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||||||||   ||||||||||||||||||   ||||||||||||||||||:::|||
Db        381 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 400

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              ||||||||||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 420

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
              |||   ||||||   :::|||      :::||||||:::      |||:::||||||
Db        421 GlnLysAlaLeuAspSerLeuLysGlyAspAsnAspAspGlnAsnMetGlyIleAsnIle 440

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::||||||:::::|||   |||:::|||||||||      ||||||||||||||
Db        441 LeuArgArgAlaIleGluSerProMetArgGlnIleValThrAsnAlaGlyTyrGluAla 460

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||:::||||||   ||||||   |||   ||||||   :::||||||||||||||:::
Db        461 SerValValValAsnLysValAlaGluHisLysAspAsnTyrGlyPheAsnAlaAlaThr 480

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||||||||   |||:::::::::   |||::::::||||||   ||||||||||||:::
Db        481 GlyGluTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgMet 500

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||   :::|||
Db        501 AlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrGluCysMetValAla 520

Qy       1567 GAAAAACCAGAACCTAAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
              :::         |||||||||:::      :::      |||   ||||||||||||||
Db        521 AspLeu------ProLysLysGluGluGlyValGlyAlaGlyAspMetGlyGlyMetGly 538

Qy       1627 GGTATGGACGGTATG 1641
              ||||||   ||||||
Db        539 GlyMetGlyGlyMet 543
```

RESULT 6
AAB69061
ID AAB69061 standard; Protein; 574 AA.
XX
AC AAB69061;
XX
DT 18-APR-2001 (first entry)
XX
DE pTrcHisB expression vector groEL fusion protein sequence.
XX
KW Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW diagnosis; detection; infection; immune response.
XX
OS Pseudomonas aeruginosa.
OS Synthetic.
XX
PN WO200102577-A1.
XX
PD 11-JAN-2001.
XX
PF 03-JUL-2000; 2000WO-GB02554.
XX
PR 01-JUL-1999; 99GB-0015419.
XX
PA (PROV-) PROVALIS UK LTD.
XX
PI Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;
XX

DR WPI; 2001-080988/09.
DR N-PSDB; AAF32453.
XX
PT Antigenic Pseudomonas aeruginosa proteins, useful in the detection
PT and/or diagnosis of P. aeruginosa infections and for producing vaccines
PT against P. aeruginosa -
XX
PS Claim 3; Fig 26; 129pp; English.
XX
CC The present invention describes antigenic Pseudomonas aeruginosa
CC proteins (P1). The P. aeruginosa proteins have antibacterial activity
CC and can be used in vaccines and as antagonists. The proteins or their
CC fragments, or antibodies are useful in the detection and/or diagnosis
CC of P. aeruginosa. They are also useful for producing a vaccine and
CC inducing an immune response against P. aeruginosa infection. An agent
CC capable of antagonising, inhibiting or otherwise interfering with the
CC function or expression of P1 are useful in the manufacture of a
CC medicament for the treatment or prophylaxis of P. aeruginosa infections.
CC The present sequence represents a specifically claimed pTrcHisB
CC expression vector groEL fusion protein sequence from the present
CC invention.
XX
SQ Sequence 574 AA;

Alignment Scores:

Pred. No.:	1.57e-155	Length:	574
Score:	1819.00	Matches:	365
Percent Similarity:	80.44%	Conservative:	75
Best Local Similarity:	66.73%	Mismatches:	103
Query Match:	61.29%	Indels:	4
DB:	22	Gaps:	1

US-09-077-574A-1 (1-1647) x AAB69061 (1-574)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||||:::   |||         |||||:::|||:::      ||||||
Db         31 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 50

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   |||||:::|||||||||   ||||||||||||||||||||||||||||:::
Db         51 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 70

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||:::|||   ||||||||||||||||||||||||||||||||||
Db         71 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 90

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::|||||||||||||||||||||||||||:::||||||:::||||||   |||
Db         91 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 110

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
                 :::|||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        111 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 130

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||:::|||   |||||||||||   |||||||||   :::|||||||||||||
Db        131 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 150

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||   ||||||:::      :::|||   ::::::   ||||||      |||
Db        151 AspLysAlaThrValAlaIleValAlaGlnLeuLysGluLeuAlaLysProCysAlaAsp 170

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||   ||||||||||||||||||||||||||||||||||||   :::||||||
Db        171 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 190

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||   ||||||||||   ||||||||||||||||||||||||||||
Db        191 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 210

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||||||||      |||   ||||||||||||||:::|||||||||||||||||||||
Db        211 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 230

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||   |||:::  |||      |||||||||:::|||   :::
Db        231 SerProTyrPheValAsnLysProAspThrMetAlaAlaGluLeuAspSerProLeuLeu 250

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::|||||||||:::::::::::::::|||||||||:::|||||   |||
Db        251 LeuLeuValAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 270

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      |||||||||||||||:::|||||||||||||||||||||||||||||
Db        271 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 290

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   ::::::|||   ||||||||||||||||||
Db        291 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 310

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||||
Db        311 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 330

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::|||      ||||||   |||
Db        331 IleSerGluGluValGlyLeuSerLeuGluGlyAlaThrLeuGluHisLeuGlyAsnAla 350

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||:::|||||||||||||||||||:::||||||||||||
Db        351 LysArgValValIleAsnLysGluAsnThrThrIleIleAspGlyAlaGlyValGlnAla 370

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||:::|||||||||   |||||||||   ||||||||||||||:::||||||||
Db        371 AspIleGluAlaArgValLeuGlnIleArgLysGlnIleGluGluThrThrSerAspTyr 390

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        391 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 410

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||||
Db        411 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 430

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||:::|||||||||||||||   |||   |||
Db        431 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 450

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::::::::   ::::::   :::|||      ::::::::::::      |||:::
Db        451 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluGluGlnAsnValGlyIle 470

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 ::::::||||||::::::|||   ||||||||||||||   ||||||||||||
Db        471 AlaLeuLeuArgArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyAsp 490

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::||||||:::||||||::::::            :::||||||||||||
Db        491 GluProSerValValValAspLysValLysGlnGlySerGlyAsnTyrGlyPheAsnAla 510

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::|||   |||   |||:::|||:::   |||::::::||||||   |||||||||
Db        511 AlaThrGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 530

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   ||||||||||   ||||||||:::      |||::::::||||||||||
Db        531 ArgSerAlaLeuGlnAlaAlaAlaSerIleGlyGlyLeuMetIleThrThrGluAlaMet 550

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              :::||||||      |||   |||      |||            |||||||||
Db        551 ValAlaGluIleValGluAspLysProAlaMet----------GlyGlyMetProAsp 566

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
              ||||||||||||   ||||||
Db        567 MetGlyGlyMetGlyGlyMet 573
```

RESULT 7
AAY23915
ID AAY23915 standard; Protein; 545 AA.
XX
AC AAY23915;
XX
DT 22-SEP-1999 (first entry)
XX
DE Amino acid sequence of a heat shock protein.
XX
KW Heat shock protein; Hsp; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9935270-A1.
XX
PD 15-JUL-1999.
XX
PF 29-DEC-1998; 98WO-CA01203.
XX
PR 31-DEC-1997; 97US-0001737.
XX
PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI Mizzen L, Wisniewski J;
XX
DR WPI; 1999-430397/36.
XX
PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
PS Disclosure; Fig 10A-E; 176pp; English.
XX
CC AAY23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.
XX
SQ Sequence 545 AA;

Alignment Scores:

Pred. No.:	3.93e-155	Length:	545
Score:	1814.50	Matches:	363
Percent Similarity:	80.88%	Conservative:	77
Best Local Similarity:	66.73%	Mismatches:	103
Query Match:	61.14%	Indels:	1
DB:	20	Gaps:	1

US-09-077-574A-1 (1-1647) x AAY23915 (1-545)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::   |||      :::   |||:::|||:::      ||||||
Db          1 MetAlaAlaLysAspValGlnPheGlyAsnGluValArgGlnLysMetValAsnGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   |||||||||||||||:::|||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValArgValThrLeuGlyProLysGlyArgAsnValValVal 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::::::::||||||   |||   ||||||||||||||||:::||||||||||||||||
```

```
Db         41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db         81 ThrAsnAspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||:::|||   |||   ||||||   ||||||      :::|||||||||||||
Db        101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||:::   :::||||||   :::|||   ||||||
Db        121 AspLysAlaValAlaAlaLeuValGluGluLeuLysAsnIleAlaLysProCysAspThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||||||||||||||||||:::||||||||||||||||||      :::|||
Db        141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              |||||||||||||||||   ||||||||||||   |||||||||||||||||||:::
Db        161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||   ||||||      ||||||||||||||||||||:::||||||||||||||||||
Db        181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::   :::   ||||||   :::      ||||||||||||:::::
Db        201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::||||||||:::::::::::|||:::||||||:::||||||||||||
Db        221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||   :::||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaSerArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::|||||||   |||:::   ||||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              |||||:::|||||||||||||||||:::|||||||||||||||||||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::|||   ||||||   |||
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuAspAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||:::   |||   ||||||||||||||||:::||||||   |||   :::
Db        321 LysArgIleGluIleGlyLysGluAsnThrThrIleIleAspGlyPheGlyAspAlaAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              |||:::|||||||||   :::||||||   |||||||||      :::|||||||||
Db        341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||:::|||||||||||||||||:::|||||||||   ||||||||||||||||||
Db        361 AspLysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||:::|||   |||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuLeu 420
```

```
Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrGlyAsnAlaAspGlnAspAlaGlyVal 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValValValAsnLysValLeuGluGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 GlySerGlyGluTyrGlyAspMetIleGluMetGlyValLeuAspProAlaLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnHisAlaAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        521 IleAlaGluIleProGluGluLysProAlaMetPro---AspMetGlyGlyMetGlyGly 539

Qy       1621 ATGGGTGGTATG 1632
Db        540 MetGlyGlyMet 543
```

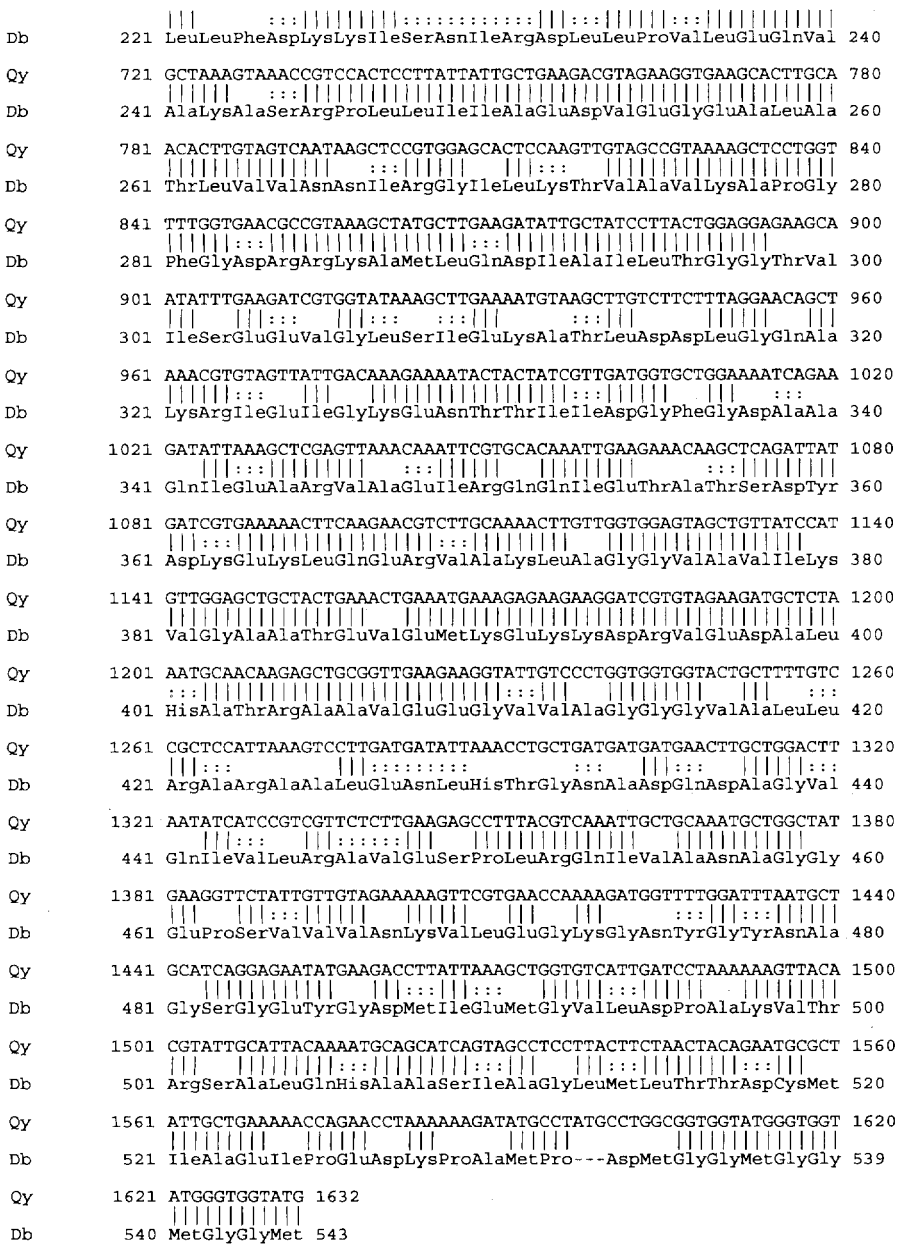
RESULT 8
AAY75747
ID AAY75747 standard; Protein; 544 AA.
XX
AC AAY75747;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 982 protein sequence SEQ ID NO:2966.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54509.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1388; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 544 AA;

Alignment Scores:

Pred. No.:	5.96e-155	Length:	544
Score:	1812.50	Matches:	362
Percent Similarity:	80.88%	Conservative:	78
Best Local Similarity:	66.54%	Mismatches:	103
Query Match:	61.07%	Indels:	1
DB:	21	Gaps:	1

US-09-077-574A-1 (1-1647) x AAY75747 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValGlnPheGlyAsnGluValArgGlnLysMetValAsnGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAsnAlaValArgValThrLeuGlyProLysGlyArgAsnValValVal 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValAlaAlaLeuValGluGluLeuLysAsnIleAlaLysProCysAspThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
```



RESULT 9
ABP77245
ID ABP77245 standard; Protein; 544 AA.
XX
AC ABP77245;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 1020.
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX
DR WPI; 2003-058415/05.
DR N-PSDB; ABZ38215.
XX
PT New protein from Neisseria gonorrheae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrheae infection -
XX
PS Disclosure; Page 262; 815pp; English.
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 544 AA;

Alignment Scores:

Pred. No.:	2.56e-154	Length:	544
Score:	1805.50	Matches:	361
Percent Similarity:	80.70%	Conservative:	78
Best Local Similarity:	66.36%	Mismatches:	104
Query Match:	60.83%	Indels:	1
DB:	24	Gaps:	1

US-09-077-574A-1 (1-1647) x ABP77245 (1-544)

```
Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValAlaAlaLeuValGluGluLeuLysAsnIleAlaLysProCysAspThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaSerArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyValVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuAspAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysArgIleGluIleGlyLysGluAsnThrThrValIleAspGlyPheGlyAspAlaAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspLysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrGlyAsnAlaAspGlnAspAlaGlyVal 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValValValAsnLysValLeuGluGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 GlySerGlyGluTyrGlyAspMetIleGlyMetGlyValLeuAspProAlaLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnHisAlaAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        521 IleAlaGluIleProGluGluLysProAlaValPro---AspMetGlyGlyMetGlyGly 539

Qy       1621 ATGGGTGGTATG 1632
Db        540 MetGlyGlyMet 543
```

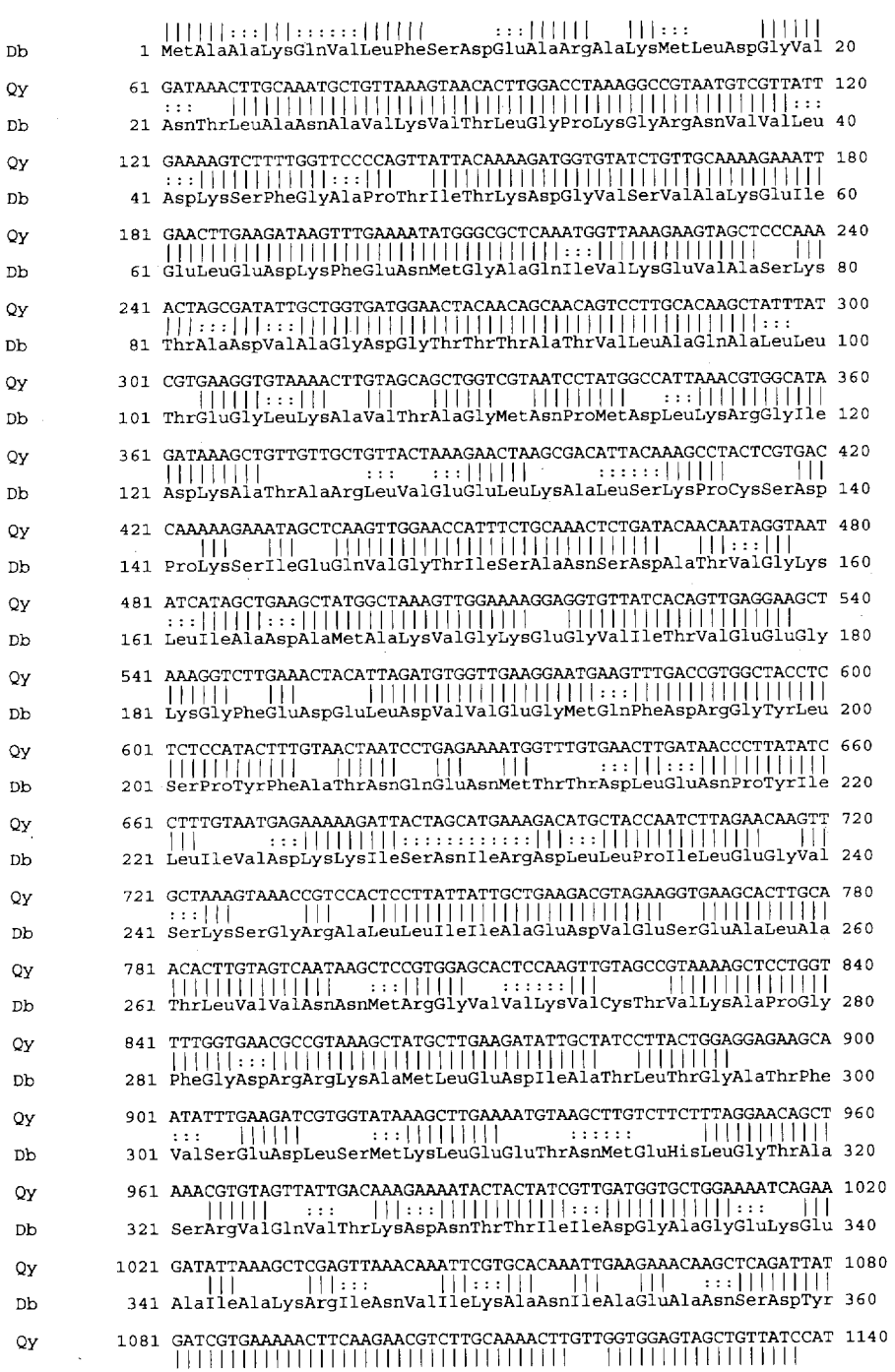
RESULT 10
AAE22070
ID AAE22070 standard; Protein; 544 AA.
XX
AC AAE22070;
XX
DT 25-JUL-2002 (first entry)
XX
DE Francisella tularensis GroEL protein.
XX
KW Heat shock protein; hsp60; GroEL protein; therapy; infection; antibiotic.
XX
OS Francisella tularensis.
XX
PN WO200218600-A1.
XX
PD 07-MAR-2002.
XX
PF 24-AUG-2001; 2001WO-GB03835.
XX
PR 26-AUG-2000; 2000GB-0021008.
XX
PA (MINA) UK SEC FOR DEFENCE.
XX
PI Titball RW, Green M, Hartley MG, Mack KA, Sjoestedt A;
XX
DR WPI; 2002-362179/39.
XX
PT Novel polypeptide comprising a component of heat shock protein, Hsp60
PT obtainable from Francisella tularensis, useful for preventing or
PT treating infection caused by Francisella tularensis in an animal -
XX
PS Disclosure; Fig 1; 26pp; English.
XX
CC The invention relates to a polypeptide comprising at least one
CC component of an heat shock protein 60 (hsp60) obtainable from
CC Francisella tularensis, and in particular the GroEL protein or its
CC fragment, and a nucleic acid sequence encoding the polypeptide. The
CC polypeptide and polynucleotides of the invention are useful for
CC preparing a medicament and for preventing or treating infection
CC caused by F. tularensis in an animal. The polynucleotide of the
CC invention is administered to an animal in a form in which it is
CC expressed in situ. The present sequence is Francisella tularensis
CC GroEL protein.
XX
SQ Sequence 544 AA;

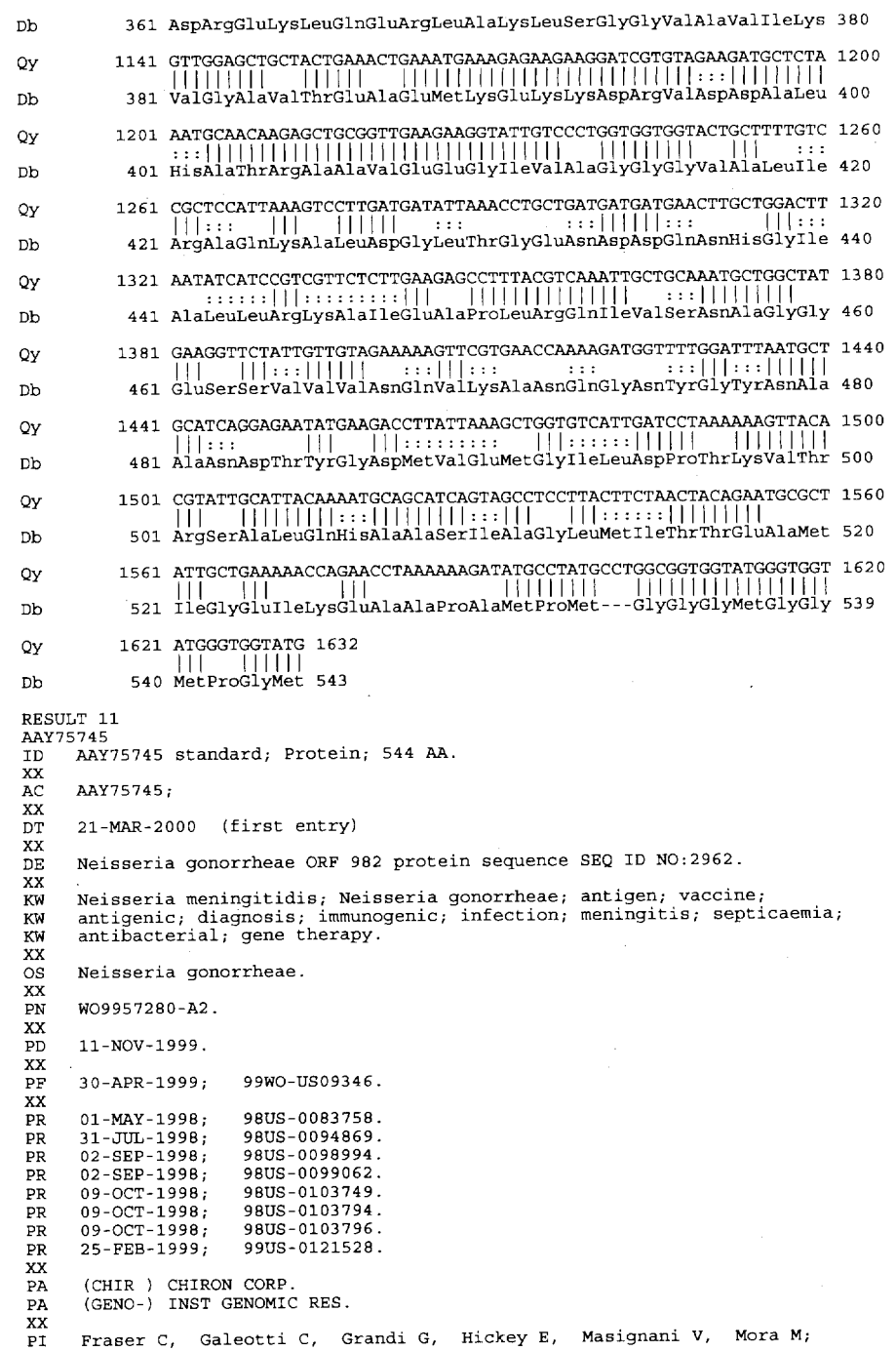
Alignment Scores:

Pred. No.:	4.59e-151	Length:	544
Score:	1769.50	Matches:	353
Percent Similarity:	79.04%	Conservative:	77
Best Local Similarity:	64.89%	Mismatches:	113
Query Match:	59.62%	Indels:	1
DB:	23	Gaps:	1

US-09-077-574A-1 (1-1647) x AAE22070 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
```





RESULT 11
AAY75745
ID AAY75745 standard; Protein; 544 AA.
XX
AC AAY75745;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 982 protein sequence SEQ ID NO:2962.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54507.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1385; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 544 AA;

Alignment Scores:

Pred. No.:	2.39e-149	Length:	544
Score:	1750.50	Matches:	353
Percent Similarity:	79.23%	Conservative:	78
Best Local Similarity:	64.89%	Mismatches:	112
Query Match:	58.98%	Indels:	1
DB:	21	Gaps:	1

US-09-077-574A-1 (1-1647) x AAY75745 (1-544)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              :::||||||:::   :::   ||||||   ::      :::|||:::      ||||||
Db          1 IleAlaSerGlnAsnLeuArgPheAspAsnArgPheLeuGlnLysMetValAsnGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::  ||      |||      |||   ||||||  ||||||||||||||||||:::
Db         21 AsnIleLeuProAlaAlaAspTrpValAlaLeuGlyAlaLysGlyArgAsnValValVal 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::::::::||||||  |||  ||||||||||||||||:::|||||||||||||||||
Db         41 AspArgAlaPheGlyGlyProHisIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||:::|||||||||||||||||||||||||||||||||||||||:::|||
Db         81 ThrAsnAspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||:::|||  |||  ||||||  ||||||      :::||||||||||||
Db        101 AlaGluGlyMetLysTyrValThrAlaGlyMetAsnProThrAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||  |||:::  :::||||||  :::|||  ||||||
Db        121 AspLysAlaValAlaAlaLeuValGluGluLeuLysAsnIleAlaLysProCysAspThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||:::|||||||||||||||||||      :::|||
Db        141 SerLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspGluGlnValGlyAla 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||  ||||||||||  ||||||||||||||||:::
Db        161 IleIleAlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||   ||||||      ||||||||||||||||||||:::||||||||||||||||||
Db        181 LysSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::  :::  ||||||  :::      ||||||||||||:::::
Db        201 SerProTyrPheIleAsnAspAlaGluLysGlnIleAlaGlyLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::||||||||:::::::::::||:::||||||:::||||||||||||
Db        221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||  :::||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaSerArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||  :::||||||  |||:::  ||||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||:::||||||||||||||||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyValVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||  |||:::  |||:::  ||||||      :::|||  ||||||
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrLeuAspAspLeuGlyGlnThr 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||:::  |||  :::||||||||||||:::::||||||  |||  :::
Db        321 LysArgIleGluIleGlyGluGluAsnThrThrValIleAspGlyPheGlyAspAlaAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                |||:::|||||||||  :::||||||  ||||||||      :::|||||||||
Db        341 GlnIleGluAlaArgValAlaGluIleArgGlnGlnIleGluThrAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||:::|||||||||||||||||:::|||||||||  ||||||||||||||||||
Db        361 AspLysGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAspArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||:::|||  ||||||||  |||  ::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::          |||:::::::::          :::   |||:::   ||||||:::
Db        421 ArgAlaArgAlaAlaLeuGluAsnLeuHisThrGlyAsnAlaAspGlnAspAlaGlyVal 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                |||:::   |||::::::|||   ||||||||||||||   ||||||||||||
Db        441 GlnIleValLeuArgAlaValGluSerProLeuArgGlnIleValAlaAsnAlaGlyGly 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||  |||:::||||||  ||||||  |||  |||       :::|||:::||||||
Db        461 GluProSerValValValAsnLysValLeuGluGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
                ||||||||||||  |||:::|||      ||||||:::||||||  ||||||||
Db        481 GlySerGlyGluTyrGlyAspMetIleGlyMetGlyValLeuAspProAlaLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||  ||||||||:::||||||||:::|||  |||:::||||||||||:::|||
Db        501 ArgSerAlaLeuGlnHisAlaAlaSerIleAlaGlyLeuMetLeuThrThrAspCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              ||||||||  ||||||  |||      :::|||          ||||||||||||||
Db        521 IleAlaGluIleProGluGluLysProAlaValPro---AspMetGlyGlyMetGlyGly 539

Qy       1621 ATGGGTGGTATG 1632
```

```
Db        540 MetGlyGlyMet 543

RESULT 12
ABB99014
ID   ABB99014 standard; protein; 545 AA.
XX
AC   ABB99014;
XX
DT   12-DEC-2002  (first entry)
XX
DE   Detrimental organism controller protein #6.
XX
KW   Detrimental organism; enterobacter.
XX
OS   Enterobacter aerogenes.
XX
PN   JP2002223764-A.
XX
PD   13-AUG-2002.
XX
PF   31-JAN-2001; 2001JP-0023255.
XX
PR   31-JAN-2001; 2001JP-0023255.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR   WPI; 2002-694867/75.
DR   N-PSDB; ABQ81658.
XX
PT   A novel detrimental organism controller for limiting detrimental
PT   organisms -
XX
PS   Claim 8; Page 41-43; 78pp; Japanese.
XX
CC   The invention relates to a protein that acts as a detrimental organism
CC   controller and is isolated from Enterobacter aerogenes. The invention
CC   encompasses a detrimental organism-resistant plant producing the protein
CC   of the invention in an effective amount for controlling detrimental
CC   organisms. A composition of the invention is useful for the preparation
CC   of a detrimental organism controller. The current sequence represents a
CC   detrimental organism controller protein of the invention.
XX
SQ   Sequence   545 AA;

Alignment Scores:
Pred. No.:                3.27e-149      Length:          545
Score:                    1749.00        Matches:         354
Percent Similarity:       78.21%         Conservative:    73
Best Local Similarity:    64.84%         Mismatches:      115
Query Match:              58.93%         Indels:          4
DB:                       23             Gaps:            2

US-09-077-574A-1 (1-1647) x ABB99014 (1-545)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300

Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValValAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyAspGluAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AlaIleGlnGlyArgValThrGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgValAlaSerLysIleAlaAspLeuLysGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460
```

```
Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValValAlaAsnThrValLysAlaGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614
Db        521 ValThrAspLeu------ProLysSerAspAlaProAspLeuGlyAlaAlaGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATG 1632
Db        539 GlyGlyMetGlyGlyMet 544
```

RESULT 13
AAB19080
ID AAB19080 standard; protein; 545 AA.
XX
AC AAB19080;
XX
DT 08-FEB-2001 (first entry)
XX
DE Amino acid sequence of a 60 kDa protein from Campylobacter jejuni.
XX
KW Antigenic protein; flagellaless mutant; vaccine; poultry; Campylobacter.
XX
OS Campylobacter jejuni.
XX
PN EP1043029-A1.
XX
PD 11-OCT-2000.
XX
PF 03-APR-2000; 2000EP-0201203.
XX
PR 09-APR-1999; 99EP-0201086.
XX
PA (ALKU) AKZO NOBEL NV.
XX
PI Jacobs AAC, Van Den Bosch JF, Nuijten PJM;
XX
DR WPI; 2000-589147/56.
XX
PT Novel Campylobacter antigenic proteins, useful for the production of
PT vaccines for protecting poultry form Campylobacter infection -
XX
PS Disclosure; Page 12-14; 21pp; English.
XX
CC The present sequence represents a 60 kDa protein of Campylobacter
CC jejuni. The protein is antigenic, and is visible on a Western blot
CC of C. jejuni protein after incubation of the blot with antibodies
CC against a flagellaless mutant of C. jejuni but not visible after
CC incubation with antibodies against wild type C. jejuni. The proteins
CC are used to produce vaccines which are useful for protecting animals,
CC especially poultry, against Campylobacter, especially C. jejuni.
XX
SQ Sequence 545 AA;

Alignment Scores:

Pred. No.:	6.11e-149	Length:	545
Score:	1746.00	Matches:	353
Percent Similarity:	79.16%	Conservative:	80
Best Local Similarity:	64.53%	Mismatches:	108
Query Match:	58.83%	Indels:	6
DB:	21	Gaps:	4

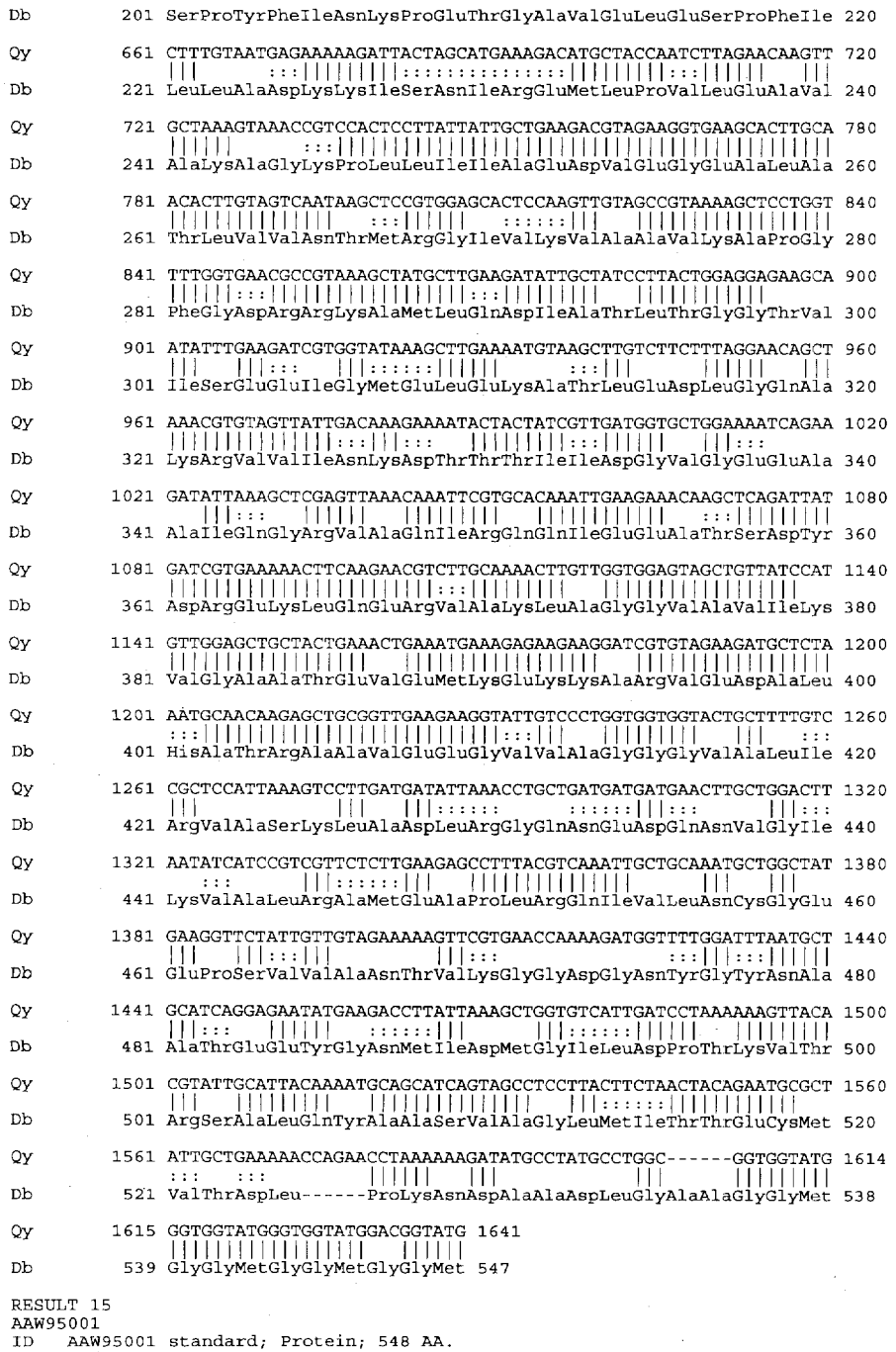
US-09-077-574A-1 (1-1647) x AAB19080 (1-545)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db          2 AlaLysGluIleIlePheSerAspGluAlaArgAsnLysLeuTyrGluGlyValLysLys 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db         22 LeuAsnAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db         42 SerPheGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluValGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db         62 LysAspSerLeuGluAsnMetGlyAlaSerLeuValArgGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db         82 AspGlnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaHisAlaIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        122 AlaCysGluAlaIleValAlaGluLeuLysLysLeuSerArgGluValLysAspLysLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        142 GluIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluLysIleGlyAsnLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluProLysSer 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db        182 IleAsnAspGluLeuAsnValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        202 TyrPheIleThrAsnAlaGluLysMetThrValGluLeuSerSerProTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        222 PheAspLysLysIleThrAsnLeuLysAspLeuLeuProValLeuGluGlnIleGlnLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        242 ThrGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleSerAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db        282 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db        302 GluGluLeuGlyArgThrLeuGluSerAlaThrIleGlnAspLeuGlyGlnAlaSerSer 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db        322 ValIleIleAspLysAspAsnThrThrIleValAsnGlyAlaGlyGluLysAlaAsnIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db        342 AspAlaArgValAsnGlnIleLysAlaGlnIleAlaGluThrThrSerAspTyrAspArg 361
```

```
Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db        382 AlaThrThrGluThrGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleLysAla 421

Qy       1267 ---ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAAT 1323
Db        422 LysAlaLysIleLysLeuAspLeuGln---------GlyAspGluAlaIleGlyAlaAla 438

Qy       1324 ATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAA 1383
Db        439 IleValGluArgAlaLeuArgAlaProLeuArgGlnIleAlaGluAsnAlaGlyPheAsp 458

Qy       1384 GGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAATGCT 1440
Db        459 AlaGlyValValValAsnSerValGluAsnAlaLysAspGluAsnThrGlyPheAspAla 478

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        479 AlaLysGlyGluTyrValAsnMetLeuGluSerGlyIleIleAspProValLysValGlu 498

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        499 ArgValAlaLeuLeuAsnAlaValSerValAlaSerMetLeuLeuThrThrGluAlaThr 518

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        519 IleSerGluIleLysGluAspLysProThrMetPro---AspMetSerGlyMetGlyGly 537

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
Db        538 MetGlyGlyMetGlyGlyMet 544
```

```
RESULT 14
AAW65070
ID   AAW65070 standard; Protein; 548 AA.
XX
AC   AAW65070;
XX
DT   11-SEP-1998  (first entry)
XX
DE   E. coli GroEL protein.
XX
KW   Chaperone; GroEL; apical region; disease; protein structure;
KW   Alzheimer's disease; reconditioning; purification; inactivation;
KW   denaturation; therapeutic.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..548
FT                   /label= GroEL
XX
PN   WO9813496-A1.
XX
PD   02-APR-1998.
XX
PF   26-SEP-1997;   97WO-GB02652.
XX
PR   03-DEC-1996;   96WO-GB02980.
PR   26-SEP-1996;   96GB-0020243.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Altamirano MM,  Fersht AR,  Zahn R;
XX
DR   WPI; 1998-230706/20.
XX
PT   New chaperone polypeptide(s) derived from GroEL - useful for, e.g.
PT   refolding of proteins and for treating diseases associated with
PT   protein structure such as Alzheimer's disease
XX
PS   Claim 1; Fig 7; 94pp; English.
XX
CC   This sequence represents the E. coli chaperone protein GroEL which can be
CC   used for treating diseases associated with protein/polypeptide structure,
CC   e.g. Alzheimer's disease. The sequence can also be used for
CC   reconditioning molecules, especially a protein that has previously been
CC   inactivated or denatured, or more generally for altering the structure of
CC   a molecule by unfolding or refolding. The sequence also has applications
CC   for purification procedures and can increase yield, specific activity or
CC   quality of biological molecules, especially by co-expression with the
CC   molecules. Therapeutic agents are administered orally, by injection or
CC   intranasally, or from a slow-release formulation.
XX
SQ   Sequence   548 AA;
```

```
Alignment Scores:
Pred. No.:                6.12e-149      Length:          548
Score:                    1746.00        Matches:         354
Percent Similarity:       77.96%         Conservative:    74
Best Local Similarity:    64.48%         Mismatches:      117
Query Match:              58.83%         Indels:          4
DB:                       19             Gaps:            2

US-09-077-574A-1 (1-1647) x AAW65070 (1-548)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
```



```
Db        201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::|||||||||:::::::::::::::::|||||||||:::||||||   |||
Db        221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      :::||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   ::::::|||   |||||||||||||||||
Db        261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||:::|||||||||   ||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||::::::||||||      :::|||      ||||||   |||
Db        301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||:::|||:::   |||||||||:::||||||   |||:::
Db        321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||:::   ||||||   |||||||||   ||||||||||||   :::|||||||||
Db        341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||:::|||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||   ||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||:::|||   |||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||              |||   |||::::::      ::::::|||:::      |||:::
Db        421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 :::      |||::::::|||   ||||||||||||||      |||   |||
Db        441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::|||        |||:::                :::|||:::||||||
Db        461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::   ||||||   :::::|||      |||::::::||||||   |||||||||
Db        481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   ||||||||   ||||||||||||||   |||::::::||||||||||
Db        501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614
              :::   :::         ||||||   |||              |||   |||||||||
Db        521 ValThrAspLeu------ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
              ||||||||||||||||   ||||||
Db        539 GlyGlyMetGlyGlyMetGlyGlyMet 547
```

RESULT 15
AAW95001
ID AAW95001 standard; Protein; 548 AA.

XX
AC AAW95001;
XX
DT 12-MAY-1999 (first entry)
XX
DE WO9902989 Seq ID 10.
XX
KW Chaperone; molecular chaperone; groEL gene.
XX
OS Escherichia coli.
XX
PN WO9902989-A1.
XX
PD 21-JAN-1999.
XX
PF 10-JUL-1998; 98WO-GB02042.
XX
PR 28-AUG-1997; 97GB-0018261.
PR 10-JUL-1997; 97GB-0014582.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Chatellier J, Fersht A;
XX
DR WPI; 1999-121113/10.
XX
PT In vivo assay for the detection of chaperone fragment activity -
PT useful for, e.g. complementing a molecular chaperone defect in vivo
XX
PS Disclosure; Page 52-55; 64pp; English.
XX
CC The invention provides a method of providing chaperone activity in vivo,
CC by administering to a cell a fragment of a molecular chaperone which has
CC in vivo activity. Also provided is a method of determining whether a
CC fragment of a molecular chaperone is active in vivo by: (i) providing a
CC cell with a deficient molecular chaperone activity; (ii) administering
CC the fragment to the cell; and (iii) determining whether the fragment
CC complements the deficient endogenous molecular chaperone. The method can
CC determine which fragments of the molecular chaperone can be used to
CC complement a mutant or deficient molecular chaperone in vivo. The method
CC allows an in vivo assessment of chaperone activity as compared to
CC previous tests which only measured activity in vitro.
XX
SQ Sequence 548 AA;

Alignment Scores:

Pred. No.:	6.12e-149	Length:	548
Score:	1746.00	Matches:	354
Percent Similarity:	77.96%	Conservative:	74
Best Local Similarity:	64.48%	Mismatches:	117
Query Match:	58.83%	Indels:	4
DB:	20	Gaps:	2

US-09-077-574A-1 (1-1647) x AAW95001 (1-548)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::   |||         ||||||   |||:::   |||||||||
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||:::|||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||::::|||   |||||||||||||||||||||||||:::||||||
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
                 :::|||   |||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360

Db       101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db       121 AspLysAlaValThrAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db       141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db       161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db       181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db       201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db       221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db       241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db       261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db       281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db       301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db       321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyGluGluAla 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db       341 AlaIleGlnGlyArgValAlaGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db       361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db       381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db       401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db       421 ArgValAlaSerLysLeuAlaAspLeuArgGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db       441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db       461 GluProSerValValAlaAsnThrValLysGlyGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy      1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db       481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy      1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db       501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy      1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614

Db       521 ValThrAspLeu------ProLysAsnAspAlaAlaAspLeuGlyAlaAlaGlyGlyMet 538

Qy      1615 GGTGGTATGGGTGGTATGGACGGTATG 1641

Db       539 GlyGlyMetGlyGlyMetGlyGlyMet 547
```

Search completed: January 28, 2004, 13:25:21
Job time : 79 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 12:47:27 ; Search time 2467 Seconds
(without alignments)
16225.972 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 1647
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	571.8	34.7	2265	11	AY109623	AY109623 Zea mays
	2	514.4	31.2	2409	11	BC047350	BC047350 Homo sapi
	3	493.6	30.0	2246	11	AK088844	AK088844 Mus muscu
	4	469.2	28.5	2290	11	AY104969	AY104969 Zea mays
	5	460.2	27.9	2338	11	AY103778	AY103778 Zea mays
	6	394.4	23.9	1376	11	AY106465	AY106465 Zea mays
	7	394.4	23.9	2143	11	AY108560	AY108560 Zea mays
	8	350.4	21.3	1201	9	AL559862	AL559862 AL559862
	9	323.4	19.6	1090	12	BM799922	BM799922 AGENCOURT
c	10	321	19.5	1327	28	BH770613	BH770613 LLMGtag37
	11	320.6	19.5	1387	11	AY108518	AY108518 Zea mays
	12	318.8	19.4	1172	9	AL515262	AL515262 AL515262
	13	314.8	19.1	1201	9	AL518632	AL518632 AL518632
	14	312.6	19.0	1201	13	BX439563	BX439563 BX439563
	15	312.2	19.0	878	10	BF275584	BF275584 GA__Eb002
	16	311.2	18.9	810	14	CD445821	CD445821 EL01T0203
	17	311	18.9	1201	9	AL557150	AL557150 AL557150
	18	310.4	18.8	1201	9	AL515579	AL515579 AL515579
	19	308.8	18.7	1061	9	AL532233	AL532233 AL532233
	20	305.6	18.6	1201	13	BX446493	BX446493 BX446493
	21	304.8	18.5	1201	13	BX462128	BX462128 BX462128
	22	303.4	18.4	1201	9	AL557181	AL557181 AL557181
	23	299	18.2	1119	12	BM458421	BM458421 AGENCOURT
	24	297.6	18.1	746	13	BU003474	BU003474 QGG35F17.
	25	296.6	18.0	815	14	CB976820	CB976820 CAB40003_
	26	296	18.0	1201	9	AL514684	AL514684 AL514684
	27	295.4	17.9	1051	12	BM478208	BM478208 AGENCOURT
	28	293.2	17.8	946	13	BQ882339	BQ882339 AGENCOURT
	29	288.4	17.5	1136	13	BU507898	BU507898 AGENCOURT
	30	288.2	17.5	1016	12	BM450130	BM450130 AGENCOURT
	31	287.8	17.5	804	14	CD434016	CD434016 EL01N0317
	32	286.8	17.4	790	10	BG599254	BG599254 EST504154
	33	286.6	17.4	1201	13	BX460128	BX460128 BX460128
	34	286.2	17.4	1086	13	BX404534	BX404534 BX404534
	35	286	17.4	876	13	BQ231270	BQ231270 AGENCOURT
	36	285.2	17.3	1016	12	BM472194	BM472194 AGENCOURT
	37	284.8	17.3	977	10	BG321293	BG321293 Zm04_04g0
	38	284.4	17.3	1055	12	BM471035	BM471035 AGENCOURT
	39	283.4	17.2	895	13	BQ216655	BQ216655 AGENCOURT
	40	281.6	17.1	811	14	CB656065	CB656065 OSJNEc09P
	41	280.8	17.0	864	14	CB634737	CB634737 OSIIEb14H
	42	280.4	17.0	850	14	CA792748	CA792748 AGENCOURT
	43	279.8	17.0	744	12	BI176450	BI176450 EST517395
	44	279.8	17.0	1201	13	BX421838	BX421838 BX421838
	45	279.8	17.0	1201	13	BX425282	BX425282 BX425282

ALIGNMENTS

RESULT 1
AY109623
LOCUS AY109623 2265 bp mRNA linear HTC 17-OCT-2002
DEFINITION Zea mays CL2221_1 mRNA sequence.
ACCESSION AY109623
VERSION AY109623.1 GI:21213414
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2265)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M.; Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2265)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers
source 1. .2265
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630867"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 615 a 430 c 599 g 538 t 83 others
ORIGIN

Query Match 34.7%; Score 571.8; DB 11; Length 2265;
Best Local Similarity 59.3%; Pred. No. 2.3e-123;
Matches 964; Conservative 0; Mismatches 659; Indels 3; Gaps 1;

```
Qy          4 GCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGAT 63
Db        290 GCTGCGAAGGACATCAAGTTTGGTGTTGAGGCCCGTGCTCTGATGNNNNNNNNNNNNNNN 349
Qy         64 AAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAA 123
Db        350 NNNNNNNCAGATGCTGTCAAAGTGACAATGGGACCTAAGGGGCGCAATGTGGTTATTGAG 409
Qy        124 AAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAA 183
Db        410 CAAAGCTTTGGTGCACCGAAAGTCACAAAGGATGGTGTGACTGTAGCAAAGAGCATTGAA 469
Qy        184 CTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACT 243
Db        470 TTTAAGGATAGAGTAAAGAATGTTGGTGCAAGCCTTGTGAAACAGGTTGCTAATGCAACT 529
Qy        244 AGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGT 303
Db        530 AATGACACTGCTGGCGATGGTACCACATGTGCCACTGTTTTGACAAAAGCAATATTTACT 589
Qy        304 GAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGAT 363
Db        590 GAGGGGTGCAAATCTGTTGCGGCTGGAATGAATGCAATGGATTTAAGGCGTGGAATCTCA 649
Qy        364 AAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAA 423
Db        650 ATGGCTGTTGACGCTGTTGTGACCAATCTGAAAGGCATGGCCAGAATGATCAGCACTTCA 709
Qy        424 AAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATC 483
Db        710 GAAGAAATTGCACAGGTGGGTACAATATCAGCAAATGGGGAAAGGGAAATTGGTGAGCTT 769
Qy        484 ATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAA 543
Db        770 ATTGCCAAGGCTATGGAGAAGGTTGGCAAAGAGGGTGTCATCACTATTGCGGACGGTAAC 829
Qy        544 GGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCT 603
Db        830 ACCCTTTATAATGAGCTTGAAGTTGTGGAGGGTATGAAACTAGACAGAGGTTACATCTCT 889
Qy        604 CCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTT 663
Db        890 CCGTACTTCATTACCAACTCAAAAGCCCAGAAATGTGAATTGGAAGACCCATTGATCTTA 949
Qy        664 TGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCT 723
Db        950 ATTCATGACAAGAAGGTTACGAATATGCATGCTGTGGTTAAGGTTTTAGAAATGGCTCTG 1009
Qy        724 AAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACA 783
```

```
Db        1010 AAGAAACAAAAGCCTCTACTGATTGTTGCAGAAGATGTGGAAAGTGAAGCATTGGGCACT 1069
Qy         784 CTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTT 843
Db        1070 TTGATTATTAACAAGCTTCGTGCAGGCATCAAGGTCTGTGCTGTCAAAGCTCCTGGTTTT 1129
Qy         844 GGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATA 903
Db        1130 GGGGAAAACAGGAAGGCAAACTTACAGGACCTTGCAATCCTTACTGGAGGAGAGGTAATA 1189
Qy         904 TTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAA 963
Db        1190 ACTGAAGAACTAGGAATGAACCTTGAGAATTTTGAGCCTCACATGCTGGGTACATGCNNN 1249
Qy         964 CGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGAT 1023
Db        1250 NNGGTGACTGTCTCTAAGGATGACACTGTTATTCTTGATGGAGCCGGAGACAAGAAGTCC 1309
Qy        1024 ATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGAT 1083
Db        1310 ATTGAAGAGAGGGCAGAGCAGATTAGATCAGCAATTGAGAATAGCACTTCAGATTATGAT 1369
Qy        1084 CGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTT 1143
Db        1370 AAGGAAAAGCTCCAGGAGCGGTTGGCAAAGCTCTCTGGAGGTGTTGCTGTTCTGAAGATT 1429
Qy        1144 GGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAAT 1203
Db        1430 GGAGGAGCCAGCGAAGCAGAAGTTGGTGAGAAGAAGGATAGAGTGACAGATGCACTGAAT 1489
Qy        1204 GCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGC 1263
Db        1490 GCTACTAAAGCTGCTGTTGAAGAGGGTATTGTACCAGGTGGTGGTGTTGCTCTTCTCTAT 1549
Qy        1264 TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAAT 1323
Db        1550 GCATCGAAGGAGCTTGATAAATTGCAGACAGCGAACTTCGATCAGAAGATTGGTGTGCAA 1609
Qy        1324 ATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAA 1383
Db        1610 ATCATTCAGAATGCTTTGAAGACACCTGTACACACAATTGCCTCAAATGCTGGGGTGGAG 1669
Qy        1384 GGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGA---TGGTTTTGGATTTAATGCT 1440
Db        1670 GGAGCAGTAGTTGTAGGAAAGCTTTTGGAGCAAGAAAATACTGACCTGGGTTATGATGCT 1729
Qy        1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        1730 GCTAAAGGTGAATATGTTGACATGGTGAAGACCGGTATCATTGACCCGCTAAAGGTTATC 1789
Qy        1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        1790 AGAACTGCTTTGGTGGATGCTGCTAGTGTGTCGTCCCTGATGACAACCACGGAATCCATA 1849
Qy        1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        1850 ATTGTAGAGATTCCCAAGGAAGAGGCACCGGCTCCAGCAATGGGTGGCGGCATGGGTGGC 1909
Qy        1621 ATGGGT 1626
Db        1910 ATGGAT 1915
```

RESULT 2
BC047350
LOCUS BC047350 2409 bp mRNA linear HTC 28-FEB-2003
DEFINITION Homo sapiens, clone IMAGE:4214709, mRNA.
ACCESSION BC047350
VERSION BC047350.1 GI:28611161
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2409)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 26 Row: h Column: 17
This clone has the following problem: no 5' EST match.

FEATURES Location/Qualifiers
source 1. .2409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4214709"
/tissue_type="Brain, anaplastic oligodendroglioma with 1p/19q loss"
/clone_lib="NCI_CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 774 a 420 c 559 g 656 t
ORIGIN

Query Match 31.2%; Score 514.4; DB 11; Length 2409;
Best Local Similarity 57.5%; Pred. No. 7.5e-110;
Matches 945; Conservative 0; Mismatches 696; Indels 3; Gaps 1;

```
Qy        6 TTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAA 65
Db      285 TGCCAAAGATGTAAAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 344

Qy       66 ACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAA 125
Db      345 TTTAGCCGATGCTGTGGCCGTTACAATGGGGCCAAAGGGAAGAACAGTGATTATTGAGCA 404

Qy      126 GTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACT 185
Db      405 GAGTTGGGGAAGTCCCAAAGTAACAAAAGATGGTGTGACTGTTGCAAAGTCAATTGACTT 464

Qy      186 TGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAG 245
Db      465 AAAAGATAAATACAAGAACATTGGAGCTAAACTTGTTCAAGATGTTGCCAATAACACAAA 524

Qy      246 CGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGA 305
Db      525 TGAAGAAGCTGGGGATGGCACTACCACTGCTACTGTACTGGCACGCTCTATAGCCAAGGA 584

Qy      306 AGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAA 365
Db      585 AGGCTTCGAGAAGATTAGCAAAGGTGCTAATCCAGTGGAAATCAGGAGAGGTGTGATGTT 644

Qy      366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 425
Db      645 ATCTGTTGATGCTGTAATTGCTGAACTTAAAAAGCAGTCTAAACCTGTGACCACCCCTGA 704

Qy      426 AGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCAT 485
Db      705 AGAAATTGCACAGGTTGCTACGATTTCTGCAAACGGAGACAAAGAAATTGGCAATATCAT 764

Qy      486 AGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
Db      765 CTCTGATGCAATGAAAAAAGTTGGAAGAAAGGGTGTCATCACAGTAAAGGATGGAAAAAC 824

Qy      546 TCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
Db      825 ACTGAATGATGAATTAGAAATTATTGAAGGCATGAAGTTTGATCGAGGCTATATTTCTCC 884

Qy      606 ATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTG 665
Db      885 ATACTTTATTAATACATCAAAAGGTCAGAAATGTGAATTCCAGGATGCCTATGTTCTGTT 944

Qy      666 TAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAA 725
Db      945 GAGTGAAAAGAAAATTTCTAGTATCCAGTCCATTGTACCTGCTCTTGAAATTGCCAATGC 1004

Qy      726 AGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACT 785
Db     1005 TCACCGTAAGCCTTTGGTCATAATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACT 1064

Qy      786 TGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGG 845
Db     1065 CGTCTTGAATAGGCTAAAGGTTGGTCTTCAGGTTGTGGCAGTCAAGGCTCCAGGGTTTGG 1124

Qy      846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATT 905
Db     1125 TGACAATAGAAAGAACCAGCTTAAAGATATGGCTATTGCTACTGGTGGTGCAGTGTTTGG 1184

Qy      906 TGAAGATCGTGGTA---TAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAA 962
Db     1185 AGAAGAGGGATTGACCCTGAATCTTGAAGACGTTCAGCCTCATGACTTAGGAAAAGTTGG 1244

Qy      963 ACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGA 1022
Db     1245 AGAGGTCATTGTGACCAAAGACGATGCCATGCTCTTAAAAGGAAAAGGTGACAAGGCTCA 1304

Qy     1023 TATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGA 1082
Db     1305 AATTGAAAAACGTATTCAAGAAATCATTGAGCAGTTAGATGTCACAACTAGTGAATATGA 1364

Qy     1083 TCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGT 1142
Db     1365 AAAGGAAAAACTGAATGAACGGCTTGCAAAACTTTCAGATGGAGTGGCTGTGCTGAAGGT 1424

Qy     1143 TGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAA 1202
Db     1425 TGGTGGGACAAGTGATGTTGAAGTGAATGAAAAGAAAGACAGAGTTACAGATGCCCTTAA 1484

Qy     1203 TGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCG 1262
Db     1485 TGCTACAAGAGCTGCTGTTGAAGAAGGCATTGTTTTGGGAGGGGGTTGTGCCCTCCTTCG 1544

Qy     1263 CTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAA 1322
Db     1545 ATGCATTCCAGCCTTGGACTCATTGACTCCAGCTAATGAAGATCAAAAAATTGGTATAGA 1604

Qy     1323 TATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGA 1382
Db     1605 AATTATTAAAAGAACACTCAAAATTCCAGCAATGACCATTGCTAAGAATGCAGGTGTTGA 1664

Qy     1383 AGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGC 1442
Db     1665 AGGATCTTTGATAGTTGAGAAAATTATGCAAAGTTCCTCAGAAGTTGGTTATGATGCTAT 1724

Qy     1443 ATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACG 1502
Db     1725 GGCTGGAGATTTTGTGAATATGGTGGAAAAAGGAATCATTGACCCAACAAAGGTTGTGAG 1784
```

```
Qy       1503 TATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTAT 1562
              | ||| |||    ||||  ||    || |||||  |  |    ||  |||||  |    |
Db       1785 AACTGCTTTATTGGATGCTGCTGGTGTGGCCTCTCTGTTAACTACAGCAGAAGTTGTAGT 1844

Qy       1563 TGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTAT 1622
              | ||||  ||  ||     | ||  ||       |||  ||     |||         ||
Db       1845 CACAGAAATTCCTAAAGAAGAGAAGGACCCTGGAATGGGTGCAATGGGTGGAATGGGAGG 1904

Qy       1623 GGGTGGTATGGACGGTATGTACTA 1646
              |||      |  || |||| |||
Db       1905 TGGTATGGGAGGTGGCATGTTCTA 1928
```

RESULT 3
AK088844

LOCUS AK088844 2246 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430028C20 product:heat shock protein, 60 kDa, full insert sequence.
ACCESSION AK088844
VERSION AK088844.1 GI:26353953
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2246)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers
source 1. .2246
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM_DB:E430028C20"
/db_xref="taxon:10090"
/clone="E430028C20"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days neonate"
CDS 71. .1792
/note="unnamed protein product; heat shock protein, 60 kDa (MGD|MGI:96242, GB|X53584, evidence: BLASTN, 98%, match=2111)
putative"
/codon_start=1
/protein_id="BAC40607.1"
/db_xref="GI:26353954"
/translation="MLRLPTVLRQMRPVSRALAPHLTRAYAKDVKFGADARALMLQGV
DLLADAVAVTMGPKGRTVIIEQSWGSPKVTKDGVTVAKSIDLKDKYKNIGAKLVQDVA
NNTNEEAGDGTTTATVLARSIAKEGFEKISKGANPVEIRRGVMLAVDAVIAELKKQSK
PVTTPEEIAQVATISANGDKDIGNIISDAMKKVGRKGVITVKDGKTLNDELEIIEGMK
FDRGYISPYFINTSKGQKCEFQDAYVLLSEKKISSVQSIVPALEIANAHRKPLVIIAE
DVDGEALSTLVLNRLKVGLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEEGLNLNL
EDVQAHDLGKVGEVIVTKDDAMLLKGKGDEAHIEKRIQEITEQLDITTSEYEKEKLNE
RLAKLSDGVAVLKVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPA
LDSLKPANEDQKIGIEIIKRALKIPAMTIAKNAGVEGSLIVEKILQSSSEVGYDAMLG
DFVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDPGMGAMGGMGG
```
```

```
                     GMGGGMF"
     polyA_signal    2225..2230
                     /note="putative"
     polyA_site      2246
                     /note="putative"
BASE COUNT      671 a    413 c    581 g    581 t
ORIGIN

  Query Match             30.0%;  Score 493.6;  DB 11;  Length 2246;
  Best Local Similarity   56.7%;  Pred. No. 5.7e-105;
  Matches  932;  Conservative    0;  Mismatches  709;  Indels    3;  Gaps    1;

Qy          6 TTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAA 65
Db        148 TGCCAAAGATGTAAAATTTGGTGCGGACGCTCGAGCCTTAATGCTTCAAGGTGTAGACCT 207

Qy         66 ACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAA 125
Db        208 TTTAGCAGATGCTGTAGCTGTTACAATGGGGCCAAAGGGAAGAACTGTGATTATTGAACA 267

Qy        126 GTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACT 185
Db        268 GAGTTGGGGAAGTCCCAAAGTAACAAAAGATGGGGTCACTGTTGCAAAGTCAATTGATTT 327

Qy        186 TGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAG 245
Db        328 AAAGGATAAATACAAAAATATTGGAGCTAAACTTGTTCAGGACGTTGCCAATAACACAAA 387

Qy        246 CGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGA 305
Db        388 CGAAGAGGCTGGGGATGGCACCACCACTGCCACTGTTCTGGCACGATCTATTGCCAAGGA 447

Qy        306 AGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAA 365
Db        448 GGGCTTTGAGAAGATCAGCAAAGGGGCTAATCCAGTGGAAATCCGGAGAGGTGTGATGTT 507

Qy        366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 425
Db        508 GGCTGTGGATGCTGTAATTGCTGAACTTAAGAAACAGTCTAAACCTGTGACAACCCCTGA 567

Qy        426 AGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCAT 485
Db        568 AGAAATTGCTCAGGTTGCTACAATTTCTGCAAATGGAGACAAAGACATTGGGAACATCAT 627

Qy        486 AGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
Db        628 TTCTGATGCAATGAAAAAGGTTGGAAGAAAGGGTGTCATCACAGTGAAGGATGGAAAAAC 687

Qy        546 TCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
Db        688 CCTGAATGATGAGCTAGAAATTATTGAAGGCATGAAGTTTGATAGAGGATATATTTCCCC 747

Qy        606 ATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTG 665
Db        748 GTATTTTATTAACACATCAAAAGGTCAAAAGTGTGAATTCCAAGATGCCTATGTCTTGTT 807

Qy        666 TAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAA 725
Db        808 GAGTGAAAAGAAAATTTCCAGTGTTCAGTCCATTGTCCCTGCTCTTGAAATTGCTAATGC 867

Qy        726 AGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACT 785
Db        868 TCATCGGAAGCCATTGGTCATAATCGCCGAAGACGTTGACGGAGAAGCTCTAAGCACGCT 927

Qy        786 TGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGG 845
Db        928 GGTTTTGAACAGGCTAAAAGTTGGTCTTCAGGTTGTGGCAGTCAAAGCTCCAGGATTTGG 987

Qy        846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATT 905
Db        988 GGACAATAGGAAGAACCAGCTTAAAGATATGGCTATTGCTACTGGTGGTGCAGTGTTTGG 1047

Qy        906 TGAAGATCGTGGTA---TAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAA 962
Db       1048 AGAAGAGGGGTTGAATCTAAATCTTGAAGATGTTCAAGCTCATGACTTAGGAAAAGTTGG 1107

Qy        963 ACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGA 1022
Db       1108 GGAGGTCATTGTCACCAAAGATGATGCCATGCTTTTGAAAGGAAAAGGTGACGAAGCTCA 1167

Qy       1023 TATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGA 1082
Db       1168 CATTGAAAAACGTATTCAAGAAATCACTGAGCAGCTAGACATCACAACTAGTGAATATGA 1227

Qy       1083 TCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGT 1142
Db       1228 AAAAGAAAAGCTGAACGAGCGACTTGCTAAACTTTCAGATGGAGTAGCTGTGTTGAAGGT 1287

Qy       1143 TGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAA 1202
Db       1288 TGGAGGAACAAGTGATGTTGAAGTGAATGAGAAAAAAGACAGAGTTACTGATGCTCTCAA 1347

Qy       1203 TGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCG 1262
Db       1348 TGCTACAAGAGCAGCTGTTGAAGAAGGCATTGTTCTAGGAGGGGGCTGCGCTCTGCTTCG 1407

Qy       1263 CTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAA 1322
Db       1408 GTGCATCCCAGCCTTGGATTCATTAAAGCCTGCTAATGAAGACCAGAAAATAGGTATAGA 1467

Qy       1323 TATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGA 1382
Db       1468 AATTATTAAAAGAGCACTTAAAATTCCTGCAATGACGATTGCTAAGAATGCAGGTGTTGA 1527

Qy       1383 AGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGC 1442
Db       1528 AGGATCTTTGATAGTTGAGAAAATTCTGCAGAGTTCCTCAGAAGTTGGTTATGACGCCAT 1587

Qy       1443 ATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACG 1502
Db       1588 GCTTGGAGATTTTGTGAACATGGTGGAAAAAGGGATCATTGATCCAACAAAGGTTGTGAG 1647

Qy       1503 TATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTAT 1562
Db       1648 AACTGCCTTACTGGATGCTGCTGGGGTGGCCTCCTTGCTAACTACAGCCGAAGCTGTAGT 1707

Qy       1563 TGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTAT 1622
Db       1708 GACAGAAATTCCTAAAGAAGAGAAGGACCCTGGAATGGGTGCAATGGGTGGCATGGGAGG 1767

Qy       1623 GGGTGGTATGGACGGTATGTACTA 1646
Db       1768 GGGTATGGGAGGCGGCATGTTCTA 1791


RESULT 4
AY104969
LOCUS       AY104969               2290 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO070942 mRNA sequence.
ACCESSION   AY104969
VERSION     AY104969.1  GI:21208047
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2290)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2290)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
```

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

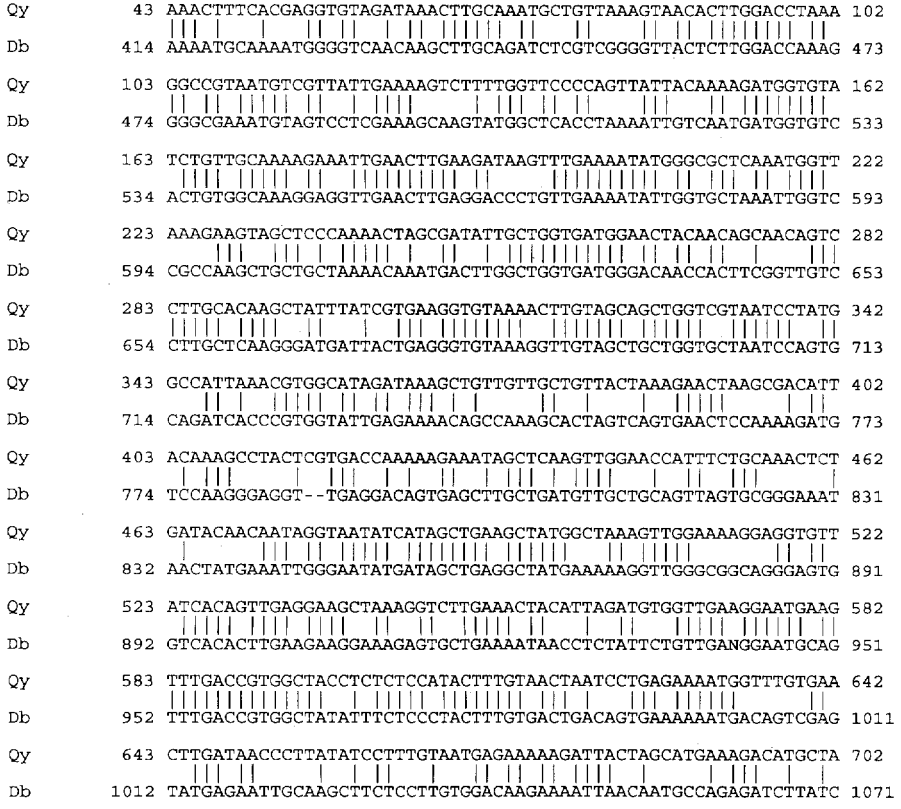
FEATURES Location/Qualifiers

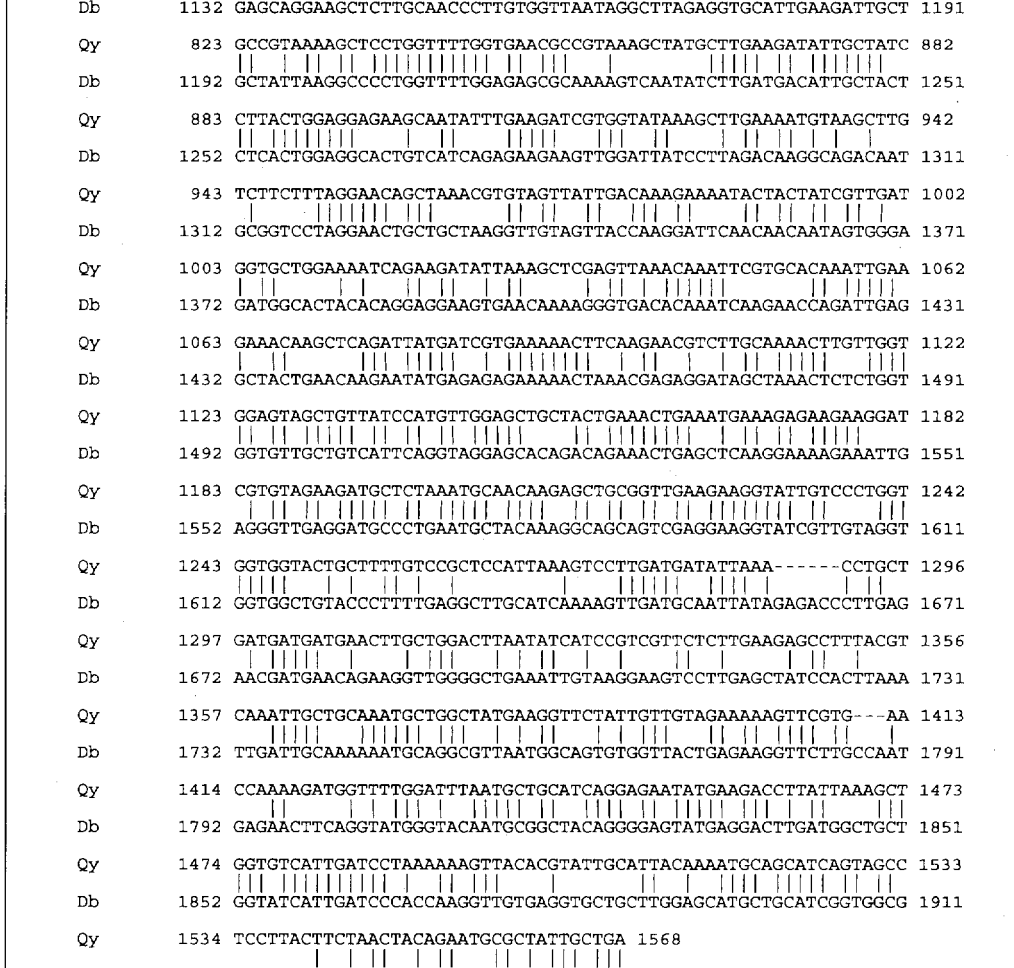
source 1. .2290
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:633867"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 684 a 457 c 583 g 565 t 1 others

ORIGIN

Query Match 28.5%; Score 469.2; DB 11; Length 2290;
Best Local Similarity 58.0%; Pred. No. 3.1e-99;
Matches 890; Conservative 0; Mismatches 634; Indels 11; Gaps 3;





RESULT 5
AY103778
LOCUS AY103778 2338 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO063180 mRNA sequence.
ACCESSION AY103778
VERSION AY103778.1 GI:21206856
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 2338)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)

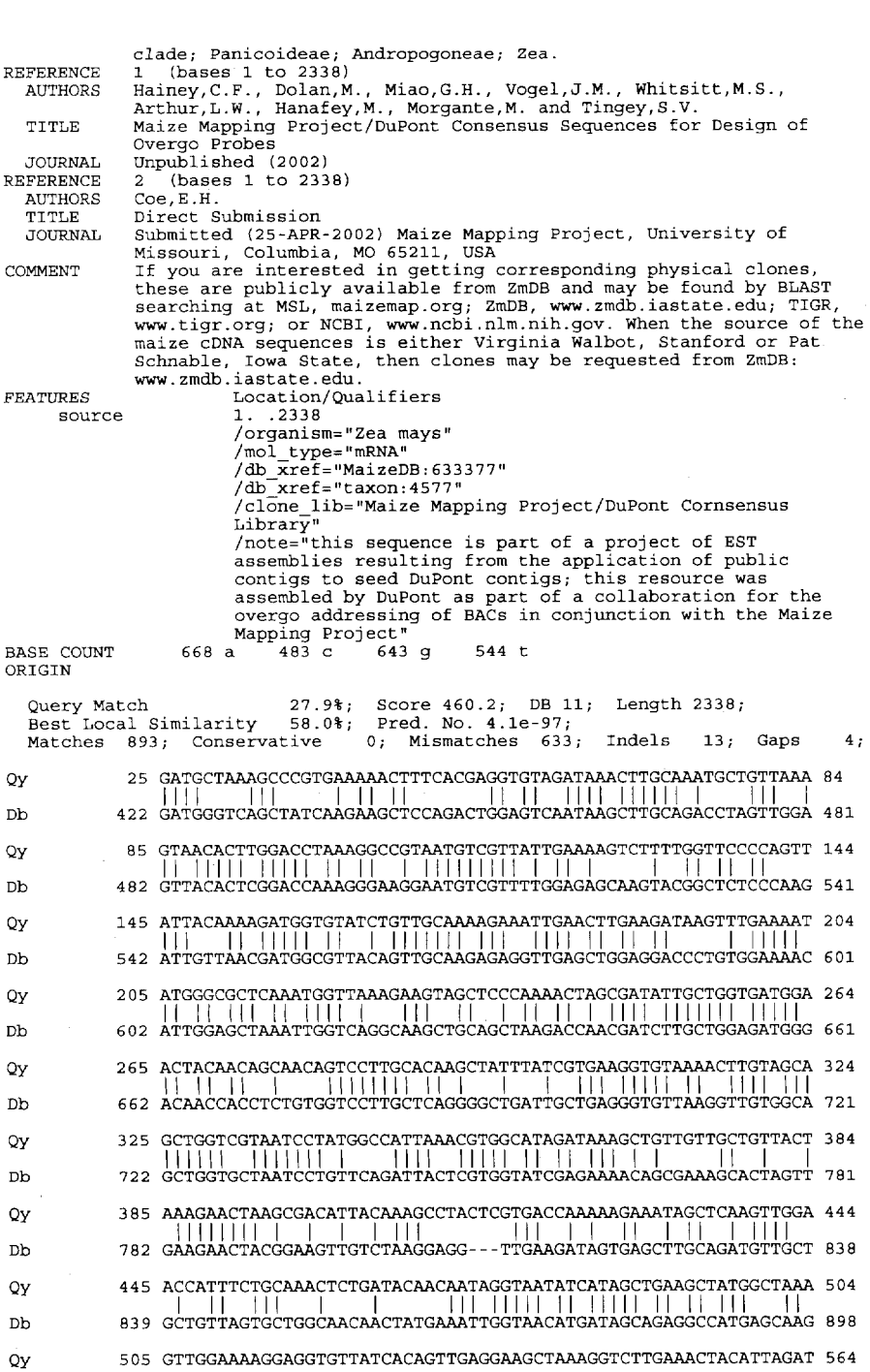
REFERENCE 2 (bases 1 to 2338)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

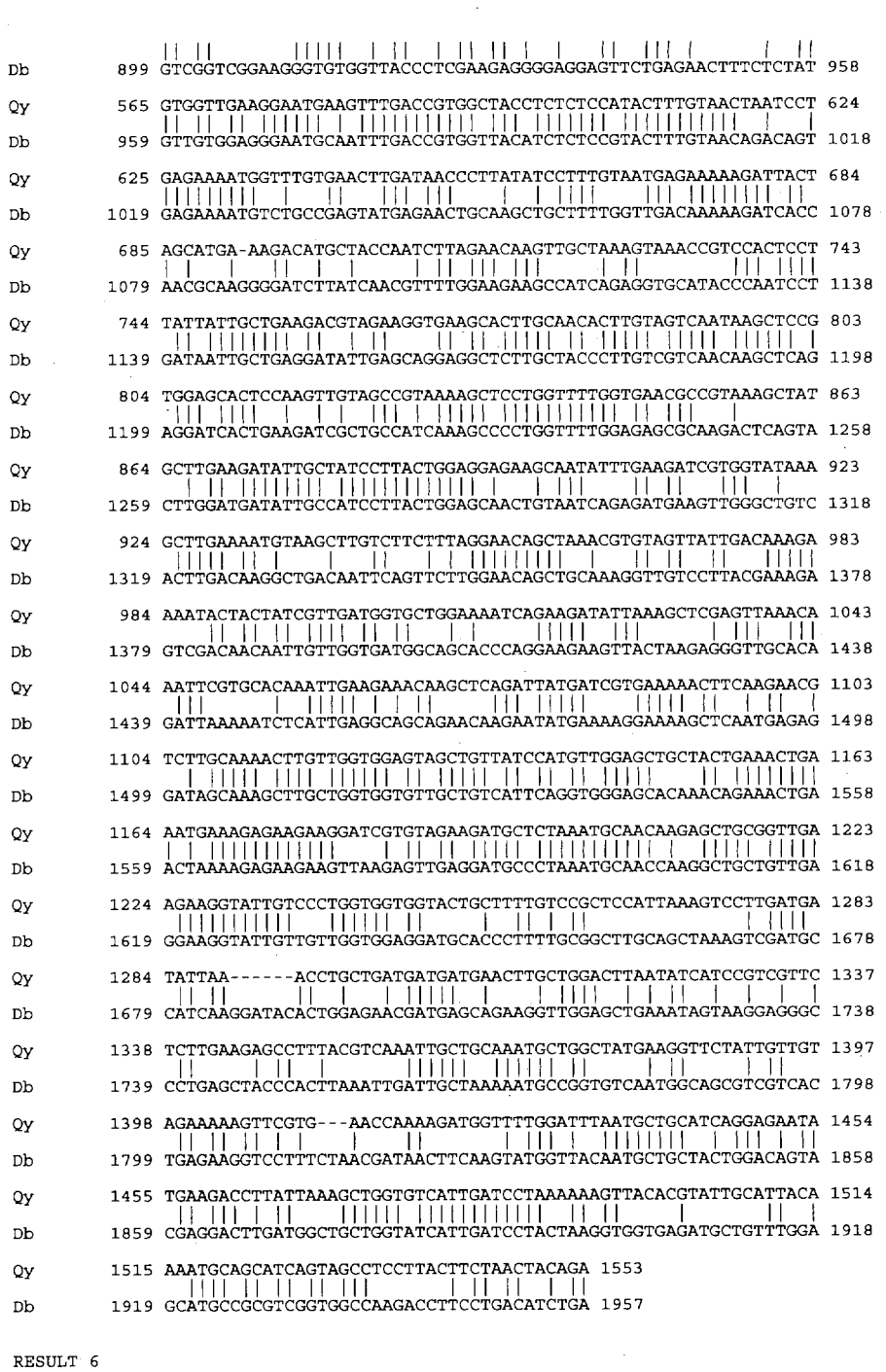
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES Location/Qualifiers
source 1. .2338
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:633377"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 668 a 483 c 643 g 544 t
ORIGIN

Query Match 27.9%; Score 460.2; DB 11; Length 2338;
Best Local Similarity 58.0%; Pred. No. 4.1e-97;
Matches 893; Conservative 0; Mismatches 633; Indels 13; Gaps 4;





RESULT 6

```
AY106465
LOCUS       AY106465                1376 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO128697 mRNA sequence.
ACCESSION   AY106465
VERSION     AY106465.1  GI:21209543
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1376)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1376)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .1376
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637486"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      397 a    277 c    359 g    343 t
ORIGIN

  Query Match           23.9%;  Score 394.4;  DB 11;  Length 1376;
  Best Local Similarity  59.0%;  Pred. No. 9.6e-82;
  Matches  696;  Conservative    0;  Mismatches  481;  Indels    3;  Gaps    1;

Qy        448 ATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAGCTGAAGCTATGGCTAAAGTT 507
              ||||||||||||  ||| |    ||| ||  || | ||| |  |||| ||||  ||||||
Db          1 ATTTCTGCAAACGGTGAAAAGGAAATTGGAGATCTAATATCAAAAGCCATGGAAAAAGTT 60

Qy        508 GGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTCTTGAAACTACATTAGATGTG 567
              |||||  |  || || ||  ||| || |  |||    | || | |    | || |
Db         61 GGAAAGGATGGAGTCATTACTATTGTTGATGGCAAAACATTGGACAATGAGCTTGAAGCA 120

Qy        568 GTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCATACTTTGTAACTAATCCTGAG 627
              ||   | |||||||||  |   | | || ||| | ||||| |||||||| ||| |||   ||
Db        121 GTACAGGGAATGAAGCTGTCCAGAGGATACATATCTCCTTACTTTGTGACTGATCAAAAG 180

Qy        628 AAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTAATGAGAAAAAGATTACTAGC 687
              |    |  |||||  | || ||||||  ||| |||    |||| || || ||  | | |
Db        181 ACTCAGAAATGTGAGATGGAGAACCCTCTTATTCTTATCCATGACAAGAAAATCTCAAAC 240

Qy        688 ATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAGTAAACCGTCCACTCCTTATT 747
              |||  |     | || |||   ||||||     |   ||      |  || || || |||
Db        241 ATGGATTCTCTCCTTCCAGCATTAGAAATTTCTATCAAGAATCGCAAGCCTCTTCTCATT 300

Qy        748 ATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTGTAGTCAATAAGCTCCGTGGA 807
              |||||||  || || |||||  || || ||  |||  |||||| | || ||||  ||||
Db        301 GTTGCTGAGGATGTTGAAGGAGATGCTCTGTCAATGCTTGTACTGAACAAGCATCGTGCT 360

Qy        808 GCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTGAACGCCGTAAAGCTATGCTT 867
              | |||| | ||    || || ||||||||||| |||||||||    | |    | ||
Db        361 GGACTCAAGGTATGTGCTGTCAAAGCTCCTGGATTTGGTGAAAATAGAAGGCACAATCTA 420

Qy        868 GAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTGAAGATCGTGGTATAAAGCTT 927
              || || || ||| |  | ||||||||||| |  ||     || || || ||| |  | |||
Db        421 GATGACATGGCTGTGATGACTGGAGGAGAGGTTATTAGCGAGGAACGCGGTCTTGATCTT 480

Qy        928 GAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTGTAGTTATTGACAAAGAAAAT 987
              |  || ||     ||          | || || ||||||   |||  | |    |  || ||
Db        481 GGCAAAGTTCAATTACAAATGCTTGGCACTGCTAAAAAGGTAACTGTATCCCTTGATGAT 540

Qy        988 ACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTAAAGCTCGAGTTAAACAAATT 1047
              ||||  ||| | |||||||  ||  |  | |  | ||  | |  |    ||||  ||
Db        541 ACTATCATCCTAGATGGTGGAGGTGACAAACAGCAGATAGATGAGAGGTGCCAACAGCTT 600

Qy       1048 CGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTGAAAAACTTCAAGAACGTCTT 1107
              | | |  ||||||   ||  |  | |  |  ||    |||||    |||| ||| |
Db        601 AGAGAATCAATTGACACAAGCACTGCTGTATTCGACAAAGAAAAGGCCCAAGAGCGTTTA 660

Qy       1108 GCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAGCTGCTACTGAAACTGAAATG 1167
              | ||  |  ||| || || ||||  || |  |||| |  |||| |||| |||| | |
Db        661 TCTAAGTTATCTGGAGGTGTTGCTGTACTCAAGATTGGTGGAGCTAGTGAAGCTGAAGTT 720

Qy       1168 AAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAACAAGAGCTGCGGTTGAAGAA 1227
              |||| || ||| | ||  |||||||||||||||  |||| |||| || || ||
Db        721 GGTGAGAAAAAAGATAGGGTGACAGATGCTCTAAATGCTGCAAGGGCTGCTGTGGAGGAG 780

Qy       1228 GGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCATTAAAGTCCTTGATGATATT 1287
              || ||||| || |||||||  |||  || |   |||  || |  |||| | |||
Db        781 GGCATTGTGCCAGGTGGTGGAGTTGCCCTTCTTTATGCCACCAAGGAGCTTGACAAGATT 840

Qy       1288 AAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCATCCGTCGTTCTCTTGAAGAG 1347
              |   | ||  | || || |||     |||| || | | || |||     |||| |  | |
Db        841 AGCACAGCAAACGAGGACGAAAAGATTGGAGTTCAAATTATCAAGAATTCTTTGAAGGCT 900

Qy       1348 CCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTTCTATTGTTGTAGAAAAAGTT 1407
              || ||       || ||||||||||||||  ||| ||  |||| ||  ||| ||  |
Db        901 CCCTTGATGACCATAGCTGCAAATGCTGGCATTGATGGAGCTATAGTCATAGGGAAGCTG 960

Qy       1408 CGTGAACCAAAAGA---TGGTTTTGGATTTAATGCTGCATCAGGAGAATATGAAGACCTT 1464
              |||||  | | ||     |||| || | | |||| |||  ||| || || | |||| |
Db        961 ATTGAACAAGAGGACCTCAGTTTGGGCTATGATGCAGCAAAAGGGGAGTACGTAGACATG 1020

Qy       1465 ATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTGCATTACAAAATGCAGCA 1524
              || || || ||  |||| |||||     || || |  ||| | ||| | || ||| |||
Db       1021 ATCAAGGCCGGCATCATCGATCCCGTGAAGGTGATCCGCACCGCACTTCAAGACGCTGCA 1080

Qy       1525 TCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTGAAAAACCAGAACCTAAA 1584
              ||  ||    |  | ||| || ||   |||| |  | ||    || | | | ||
Db       1081 AGTGTTTCCCTGCTGATGGCAACCACGGAGGCCGCTGTCTCCGAGCTCCCGGCAACCAAG 1140

Qy       1585 AAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGG 1624
                 ||  | |  | |        ||| | || ||||
Db       1141 TCCAGAATAGCCAGCCGTATGCCGCAGATGAGCGGGATGG 1180


RESULT 7
AY108560
LOCUS       AY108560                2143 bp    mRNA    linear   HTC 17-OCT-2002
DEFINITION  Zea mays PCO140434 mRNA sequence.
ACCESSION   AY108560
VERSION     AY108560.1  GI:21211654
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2143)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
```

```
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2143)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .2143
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638266"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      546 a    494 c    598 g    505 t
ORIGIN

  Query Match             23.9%;  Score 394.4;  DB 11;  Length 2143;
  Best Local Similarity   54.5%;  Pred. No. 1.2e-81;
  Matches  858;  Conservative    0;  Mismatches  706;  Indels   10;  Gaps    3;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db        275 CCAAGGAGATCGCCTTTGACCAGGGCTCCAGAGCCGCCCTTCAGGCCGGCGTCGAGAAGC 334

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db        335 TCGCCGCCGCGGTCGGCGTCACCCTCGGACCCAGAGGAAGGAATGTTGTCTTGGATGAG- 393

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        394 --TTTGGAACCCCCAAAGTGGTCAATGATGGAGTTACCATTGCCCGCGCTATTGAGCTTG 451

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        452 CTGATCCCATGGAGAATGCTGGTGCTTCCTTGATTCGTGAAGTTGCTAGCAAGACGAATG 511

Qy        248 ATATTGCTGGTGATGGAACTAC-AACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db        512 ACTCGGCTGGTGATGGAACCACAAACTGCCTCTGTTCTCGCTAGGGAGATCATCAAATTG 571

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        572 GGCATGTTGAGCATTACTTCAGGGGCAAATCCAGTGTCCGTTAAGAAGGGCATTGATAAG 631

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        632 ACTGTTCAGAAATTGGTGGAGGAACTTGAGAAGAAATCAAGGCCGGTCAAGGGTAGTGGG 691

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        692 GATATTAAAGCTGTTGCTGCCATATCAGCTGGAAATGATGAATTTGTTGGGACTATGATC 751

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        752 GCCGAAGCTATTGACAAGGTTGGCCCTGATGGTGTCCTCTCAATTGAGTCGTCATCGTCG 811

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606

Db        812 TTTGAGACCACAGTTGAAGTTGAAGAAGGGATGGAGCTTGACAGAGGATATATCTCCCCT 871

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        872 CAGTTTGTCACTAACCTTGAAAAATCTATTGTTGAGTTCGAAAATGCTCGGATTCTTGTC 931

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        932 ACTGATCAGAAGATATCATCGATAAAAGAAATTATTCCTCTGTTGGAGCAGACAACACAG 991

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        992 TTAAGAGCACCACTTCTTATAATTGCAGAGGATGTAAGTGGCGAGGCGCTGGCAACATTA 1051

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db       1052 GTCATAAACAAGCTTAGAGGAATTTTAAATGTGGCTGCAATCAAAGCTCCTGGTTTTGGT 1111

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db       1112 GAGAGGCGTAAAGCTCTTCTTCAGGACATTGCCATTGTTACAGGTGCTGAATATCAATCC 1171

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db       1172 AAAGATCTTGGTCTACTAGTTGAGGATACAACAGTGGAGCAGCTTGGCATAGCTCGGAAA 1231

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db       1232 GTTACAATCTCAAGTTCCTCGACGACCATTATAGCAGATGCTGCTAGCAAAGACGATATC 1291

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db       1292 CAGGCCAGAATCGCGCAGCTGAAGAGAGAGCTTTCTCAGACGGACTCGACGTATGACTCT 1351

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db       1352 GAGAAGTTGGCAGAGAGAATTGCAAAGCTTTCTGGTGGAGTTGCTGTGGTTAAGGTTGGA 1411

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db       1412 GCATCGACCGAGGCAGAGCTAGAGGACCGCAAGCTCCGCATAGAGGACGCGAAGAATGCG 1471

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db       1472 ACTTTTGCAGCAATAGAGGAAGGTATTGTTCCAGGAGGCGGTGCAGCATATGTTCATCTA 1531

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGC------TGATGATGATGAACTTGCTGGACTT 1320
Db       1532 TCGACATTCGTACCGGCTATCAAGGAGACGTTGGATGATCCTGAGGAGCGCCTAGGTGCT 1591

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1592 GATATCATTCAGAAGGCTTTGGTGGCACCTGCAGCGCTGATAGCCCATAATGCTGGGGTG 1651

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       1652 GAAGGTGAGGTGATTGTGGATAAAATCCGGGAAAGCGAGTGGGAGTTCGGTTACAACGCC 1711

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       1712 ATGGCGGACAAGCACGAGAACCTGGTGGAGGCTGGTGTGATCGACCCTGCCAAAGTGACT 1771

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       1772 AGGTGCGCCCTGCAGAACGCTGCCTCCGTGGCTGGAATGGTGCTGACTACCCAAGCAATC 1831

Qy       1561 ATTGCTGAAAAACC 1574
Db       1832 GTTGTGGAGAAGCC 1845

RESULT 8
```

```
AL559862
LOCUS       AL559862               1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL559862 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG003YN06 5-PRIME, mRNA sequence.
ACCESSION   AL559862
VERSION     AL559862.2  GI:31283993
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12905763.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1228.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DG003DG03QP1&cluster=1228.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DG003DG03QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG003YN06"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      369 a    200 c    291 g    302 t     39 others
ORIGIN

  Query Match             21.3%;  Score 350.4;  DB 9;  Length 1201;
  Best Local Similarity   57.4%;  Pred. No. 2e-71;
  Matches  619;  Conservative   15;  Mismatches  441;  Indels    3;  Gaps    1;

Qy        218 TGGTTAAAGAAGTAGCTCCCAAAACTAGCGATATTGCTGGTGATGGAACTACAACAGCAA 277
Db         23 TGTACAAAAAAGCAGGCTGGTACCGGTCCGGAATTCCCGGGATTGGCACTACCACTGCTA 82

Qy        278 CAGTCCTTGCACAAGCTATTTATCGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATC 337
Db         83 CTGTACTGGCACGCTCTATAGCCAAGGAAGGCTTCGAGAAGATTAGCAAAGGTGCTAATC 142

Qy        338 CTATGGCCATTAAACGTGGCATAGATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCG 397
Db        143 CAGTGGAAATCAGGAGAGGTGTGATGTTAGCTGTTGATGCTGTAATTGCTGAACTTAAAA 202

Qy        398 ACATTACAAAGCCTACTCGTGACCAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAA 457
Db        203 AGCAGTCTAAACCTGTGACCACCCCTGAAGAAATTGCACAGGTTGCTACGATTTCTGCAA 262

Qy        458 ACTCTGATACAACAATAGGTAATATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAG 517
Db        263 ACGGAGACAAAGAAATTGGCAATATCATCTCTGATGCAATGAAAAAAGTTGGAAGAAAGG 322

Qy        518 GTGTTATCACAGTTGAGGAAGCTAAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAA 577
Db        323 GTGTCATCACAGTAAAGGATGGAAAAACACTGAATGATGAATTAGAAATTATTGAAGGCA 382

Qy        578 TGAAGTTTGACCGTGGCTACCTCTCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTT 637
Db        383 TGAAGTTTGATCGAGGCTATATTTCTCCATACTTTATTAATACATCAAAAGGTCAGAAAT 442

Qy        638 GTGAACTTGATAACCCTTATATCCTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACA 697
Db        443 GTGAATTCCAGGATGCCTATGTTCTGTTGAGTGAAAAGAAAATTTCTAGTATCCAGTCCA 502

Qy        698 TGCTACCAATCTTAGAACAAGTTGCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAG 757
Db        503 TTGTACCTGCTCTTGAAATTGCCAATGCTCACCGTAAGCCTTTGGTCATAATCGCTGAAG 562

Qy        758 ACGTAGAAGGTGAAGCACTTGCAACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAG 817
Db        563 ATGTTGATGGAGAAGCTCTAAGTACACTCGTCTTGAATAGGCTAAAGGTTGGTCTTCAGG 622

Qy        818 TTGTAGCCGTAAAAGCTCCTGGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTG 877
Db        623 TTGTGGCAGTCAAGGCTCCAGGGTTTGGTGACAATAGAAAGAACCAGCTTAAAGATATGG 682

Qy        878 CTATCCTTACTGGAGGAGAAGCAATATTTGAAGATCGTGGTA---TAAAGCTTGAAAATG 934
Db        683 CTATTGCTACTGGTGGTGCAGTGTTTGGAGAAGAGGGATTGACCCTGAATCTTGAAGACG 742

Qy        935 TAAGCTTGTCTTCTTTAGGAACAGCTAAACGTGTAGTTATTGACAAAGAAAATACTACTA 994
Db        743 TTCAGCCTCATGACTTAGGAAAAGTTGGAGAGGTCATTGTGACCAAAGACGATGCCATGC 802

Qy        995 TCGTTGATGGTGCTGGAAAATCAGAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCAC 1054
Db        803 TCTTAAAAGGAAAAGGTGACAAGGCTCAAATTGAAAAACGTATTCAAGAAATCATTGAGC 862

Qy       1055 AAATTGAAGAAACAAGCTCAGATTATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAAC 1114
Db        863 AGTTAGATGTCACAACTAGTGAATATGAAAAGGAAAAACTGAATGAACGGCTTGCAAAAC 922

Qy       1115 TTGTTGGTGGAGTAGCTGTTATCCATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGA 1174
Db        923 TTTCAGATGGAGTGGCTGTGCTGAAGGTTGGTGGGGCAAGTGATGTTGAAGTGAATGAAA 982

Qy       1175 AGAAGGATCGTGTAGAAGATGCTCTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTG 1234
Db        983 ARAAAGACAGARTTACAGATGCCCTTAATGCTACAAGAGCTGCTGKTKGAGGAGGCATTG 1042

Qy       1235 TCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCATTAAAGTCCTTGATGATATTAAACC 1292
Db        1043 TTTTKKGGGGGGGKTGGKGCCTCCYTCSRTGCMTTCCMGCCCTTGRCYCWTTKACTCC 1100

RESULT 9
BM799922
LOCUS       BM799922               1090 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6417164 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492101
            5', mRNA sequence.
ACCESSION   BM799922
VERSION     BM799922.1  GI:19116745
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1090)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
          http://image.llnl.gov
          Plate: LLAM12113  row: k  column: 14
          High quality sequence stop: 743.
FEATURES             Location/Qualifiers
     source          1..1090
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5492101"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      332 a    209 c    274 g    275 t
ORIGIN

  Query Match            19.6%;  Score 323.4;  DB 12;  Length 1090;
  Best Local Similarity  59.4%;  Pred. No. 4.3e-65;
  Matches  549;  Conservative    0;  Mismatches  376;  Indels    0;  Gaps    0;

Qy          6 TTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAA 65
Db        121 TGCCAAAGATGTAAAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 180

Qy         66 ACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAA 125
Db        181 TTTAGCCGATGCTGTGGCCGTTACAATGGGGCCAAAGGGAAGAACAGTGATTATTGAGCA 240

Qy        126 GTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACT 185
Db        241 GAGTTGGGGAAGTCCCAAAGTAACAAAAGATGGTGTGACTGTTGCAAAGTCAATTGACTT 300

Qy        186 TGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAG 245
Db        301 AAAAGATAAATACAAGAACATTGGAGCTAAACTTGTTCAAGATGTTGCCAATAACACAAA 360

Qy        246 CGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGA 305
Db        361 TGAAGAAGCTGGGGATGGCACTACCACTGCTACTGTACTGGCACGCTCTATAGCCAAGGA 420

Qy        306 AGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAA 365
Db        421 AGGCTTCGAGAAGATTAGCAAAGGTGCTAATCCAGTGGAAATCAGGAGAGGTGTGATGTT 480

Qy        366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 425
Db        481 AGCTGTTGATGCTGTAATTGCTGAACTTAAAAAGCAGTCTAAACCTGTGACCACCCCTGA 540

Qy        426 AGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCAT 485
Db        541 AGAAATTGCACAGGTTGCTACGATTTCTGCAAACGGAGACAAAGAAATTGGCAATATCAT 600

Qy        486 AGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
Db        601 CTCTGATGCAATGAAAAAAGTTGGAAGAAAGGGTGTCATCACAGTAAAGGATGGAAAAAC 660

Qy        546 TCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
Db        661 ACTGAATGATGAATTAGAAATTATTGAAGGCATGAAGTTTGATCGAGGCTATATTTCTCC 720

Qy        606 ATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTG 665
Db        721 ATACTTTATTAATACATCAAAAGGTCAGAAATGTGAATTCCAGGATGCCTATGTTCTGTT 780

Qy        666 TAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAA 725
Db        781 GAGTGAAAAGAAAATTTCTAGTATCCAGTCCATTGTACCTGCTCTTGAAATTGCCCATGC 840

Qy        726 AGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACT 785
Db        841 TCACCGTAAGCCTTTGGTCATAATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACT 900

Qy        786 TGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGG 845
Db        901 CGTCTTGAATAGGGCTAAAGGTGGTCTCCAGGGTGTGGCAGTCAAGGCTCAGGGTTTTGG 960

Qy        846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATT 905
Db        961 TGACAATAGAAAGAACCAGCCTAAAGATATGGCTATTGCTACTGGTGGGGCCAAGGTTTT 1020

Qy        906 TGAAGATCGTGGTATAAAGCTTGAA 930
Db       1021 GGAAAAAAAGGGATTGACCCTTGAA 1045

RESULT 10
BH770613/c
LOCUS       BH770613                1327 bp    DNA     linear   GSS 01-MAY-2002
DEFINITION  LLMGtag372 MG1363 Random Sequence Tag Library Lactococcus lactis
            subsp. cremoris genomic, genomic survey sequence.
ACCESSION   BH770613
VERSION     BH770613.1  GI:20373570
KEYWORDS    GSS.
SOURCE      Lactococcus lactis subsp. cremoris
  ORGANISM  Lactococcus lactis subsp. cremoris
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 1327)
  AUTHORS   Bolotin,A., Ehrlich,S.D. and Sorokin,A.
  TITLE     Studies of genomes of dairy bacteria Lactococcus lactis
  JOURNAL   Sci. Aliments,  (2002) In press
COMMENT     Contact: Sorokin A
            Genetique Microbienne
            INRA
            CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
            Tel: 33 1 34 65 25 16
            Fax: 33 1 34 65 25 21
            Email: sorokine@jouy.inra.fr
            best homologue in strain IL1403 is groEL (93%)
            Class: shotgun
            High quality sequence start: 30
            High quality sequence stop: 1299.
FEATURES             Location/Qualifiers
     source          1..1327
                     /organism="Lactococcus lactis subsp. cremoris"
                     /mol_type="genomic DNA"
                     /strain="MG1363"
                     /db_xref="taxon:1359"
                     /clone_lib="MG1363 Random Sequence Tag Library"
                     /note="Vector: pSGMU2; Site_1: SmaI; Library of
                     chromosomal fragments of L.lactis strain MG1363 was
                     prepared by partial AluI digestion or by sonication."
BASE COUNT      392 a    292 c    239 g    403 t      1 others
ORIGIN

  Query Match            19.5%;  Score 321;  DB 28;  Length 1327;
  Best Local Similarity  65.7%;  Pred. No. 1.7e-64;
  Matches  498;  Conservative    0;  Mismatches  256;  Indels    4;  Gaps    2;

Qy        823 GCCGTAAAAGCTCCTGGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATC 882
Db       1321 GCTGTAAAAGCACCGGGATTTGGTG-CCGCCGTAAAGCTCAATTGGAAGATTTAGCTATC 1263

Qy        883 CTTACTGGAGGAGAAGCAATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTG 942
Db       1262 TTGACAGGTGGTACAGTAATACTGGAGGAACTTGGCTTTGATCTTAAAGATGCTACACTA 1203

Qy        943 TCTTCTTTAGGAACAGCTAAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGAT 1002
Db       1202 GAAGCTTTAGGACAAGCTGCTAAAGCATCTGTTGATAAAGACCATACAACAATTGTTGAA 1143

Qy       1003 GGTGCTGGAAAATCAGAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAA 1062
```

```
Db       1142 GGTGCTGGTTCTGTTGGTGCCATTTCTGATCGTGTGGCTATCATCAAAGCACAAATCGAA 1083

Qy       1063 GAAACAAGCTCAGATTATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGT 1122

Db       1082 AAAATAACTTCTGATTTTGCCCGTGAAAAATGNCAAGAGCGCTTAGCGAAATTAGCTGGC 1023

Qy       1123 GGAGTAGCTGTTATCCATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGAT 1182

Db       1022 GGAGTTGCGGTAGTTAAAGTAGGAGCCGCAACAGAAACAGAACTTAAAGCAATGAATCTT 963

Qy       1183 CGTGTAGAAGATGCTCTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGT 1242

Db        962 TTGATTGAAGACGCACTTAATGCAACACGTGCTGCTGTTGAAGAAGGTATCGTTTGTGGT 903

Qy       1243 GGTGGTACTGCTTTTGTCCGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGAT 1302

Db        902 GGTGGTACAGCTCTTGTTAACGCCATCGCAGCTTTGGATAAACTT---TCAGAAGAAGGC 846

Qy       1303 GATGAACTTGCTGGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATT 1362

Db        845 GACATTCAAACAGGAATTAACATCGTTCGCCGTGCCCTTGAAGACCCAGTTCGTCAAATC 786

Qy       1363 GCTGCAAATGCTGGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT 1422

Db        785 GCGGCCAATGCAGGATATGAAGGTTCAGTTATCATTGATAAACTTCGTTCAGAAAAAGTT 726

Qy       1423 GGTTTTGGATTTAATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATT 1482

Db        725 GGTACAGGATTTAATGCAGCGACAGGTCAATGGGTAAATATGATTGAAGAAGGAATCGTT 666

Qy       1483 GATCCTAAAAAAGTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTT 1542

Db        665 GACCCGGCAAAAGTGACTCGTTCAGCTTTGCAAAATGCAGCTTCTGTGGCTGGATTAATC 606

Qy       1543 CTAACTACAGAATGCGCTATTGCTGAAAAACCAGAACC 1580

Db        605 TTGACTACTGAAGCAGTTGTGGCTAATAAACCAGAACC 568


RESULT 11
AY108518
LOCUS       AY108518               1387 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO123570 mRNA sequence.
ACCESSION   AY108518
VERSION     AY108518.1  GI:21211596
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1387)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1387)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .1387
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:637188"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      426 a    257 c    372 g    332 t
ORIGIN

  Query Match             19.5%;  Score 320.6;  DB 11;  Length 1387;
  Best Local Similarity   57.8%;  Pred. No. 2.2e-64;
  Matches  611;  Conservative    0;  Mismatches  439;  Indels    7;  Gaps    2;

Qy        552 AACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCATACTT 611

Db          6 AGCCACTGTTGAAGTTGAAGAAGGAATGGAGATTGACCGCGGTTATATTTCCCCACAATT 65

Qy        612 TGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTAATGA 671

Db         66 TGTGACAAACCTCGAGAAATCAATTGTGGAGTTTGAGAATTGTAAGGTTCTTATTACTGA 125

Qy        672 GAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAGTAAA 731

Db        126 TCAGAAGATCACAAGCATAAAGGAAATCCTTCCGATTTTGGAGAAGACCACACAGCTGAG 185

Qy        732 CCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTGTAGT 791

Db        186 AGCCCCTCTGTTCATTATTGCTGAGGACATTACTGGTGAAGCTTTGGCAACTCTTGTTGT 245

Qy        792 CAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTGAACG 851

Db        246 TAACAAGCTTCGAGGAATTCTCAATGTTGCAGCGATTAAGGCTCCAAGTTTTGGTGAGCG 305

Qy        852 CCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTGAAGA 911

Db        306 GCGGAAGGCTGTACTTCAGGACATTGCCATCGTCACAGGTGCTGAGTTCCTAGCAAAAGA 365

Qy        912 TCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTGTAGT 971

Db        366 TCTTGGTTTGTTGGTTGAAAATGCTACAGAGGAACAACTTGGCACAGCAAGGAAAGTCAC 425

Qy        972 TATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTAAAGC 1031

Db        426 GATACATCAGACTACAACAACCCTCATAGCAGATGCAGCCAGTAAAGATGAGATCCAGGC 485

Qy       1032 TCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTGAAAA 1091

Db        486 GAGGATTGCACAACTGAAGAAGGAGCTTGCTGAGACTGATTCAGTTTATGATACTGAGAA 545

Qy       1092 ACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAGCTGC 1151

Db        546 ATTGGCCGAGAGAATTGCTAAGCTTGCCGGTGGTGTTGCTGTTATCAAGGTTGGAGCTGC 605

Qy       1152 TACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAACAAG 1211

Db        606 AACTGAGACTGAGCTTGAGGACCGGCAGCTAAGAATTGAGGATGCAAAAAATGCCACTTT 665

Qy       1212 AGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCATTAA 1271

Db        666 TGCAGCTATAGAGGAGGGCATTGTTCCAGGAGGTGGTGCGGCATATGTGCACCTATCTAC 725

Qy       1272 AGTCCTTGATGATATTAAACCTGCTGATGA------TGATGAACTTGCTGGACTTAATAT 1325

Db        726 GGTTGTCCCTAAGATCAAAGAAGCAATTGAGGACCCTGATGAGCGGCTTGGCGCTGATAT 785

Qy       1326 CATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGG 1385

Db        786 TATTCAAAAGGCATTGGTAGCACCTGCTTCATTGATTGCACACAATGCTGGAGTAGAGGG 845
```

```
Qy       1386 TTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGG-TTTTGGATTTAATGCTGCAT 1444
Db        846 TGAAGTGGTTGTCGAGAAGATCAAGGCAGAGCGATTGGGAGGTGGGCTACAATGCAATGA 905

Qy       1445 CAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTA 1504
Db        906 CTGACAAGTACGAGAACCTGATGGAGGCTGGTGTGATCGACCCTGCCAAGGTGACCAGGT 965

Qy       1505 TTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTG 1564
Db        966 GTGCGCTCCAGAATGCGGCCTCAGTAGCTGGAATGGTCCTAACCACGCAGGCCATTGTTG 1025

Qy       1565 CTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCC 1601
Db       1026 TTGAGAAGCCAAAGCCGAAGCCTCAAGTGGCAGAGCC 1062
```

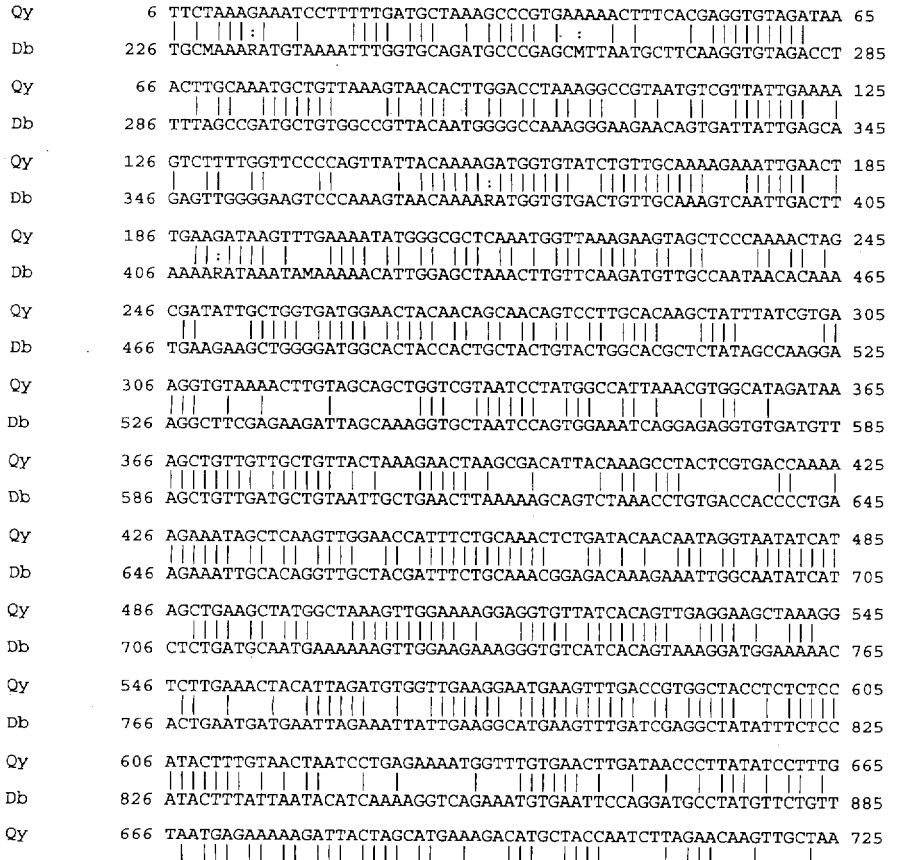
RESULT 12
AL515262
LOCUS AL515262 1172 bp mRNA linear EST 08-MAY-2003
DEFINITION AL515262 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone CL0BB016ZA04 5-PRIME, mRNA sequence.
ACCESSION AL515262
VERSION AL515262.2 GI:30465145
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1172)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12778755.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 1228.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CL0BB016ZA04RP1&cluster=1228.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CL0BB016ZA04RP1.
FEATURES Location/Qualifiers
source 1. .1172
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL0BB016ZA04"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 352 a 216 c 293 g 294 t 17 others
ORIGIN

Query Match 19.4%; Score 318.8; DB 9; Length 1172;
Best Local Similarity 59.9%; Pred. No. 5.4e-64;
Matches 543; Conservative 4; Mismatches 357; Indels 2; Gaps 1;

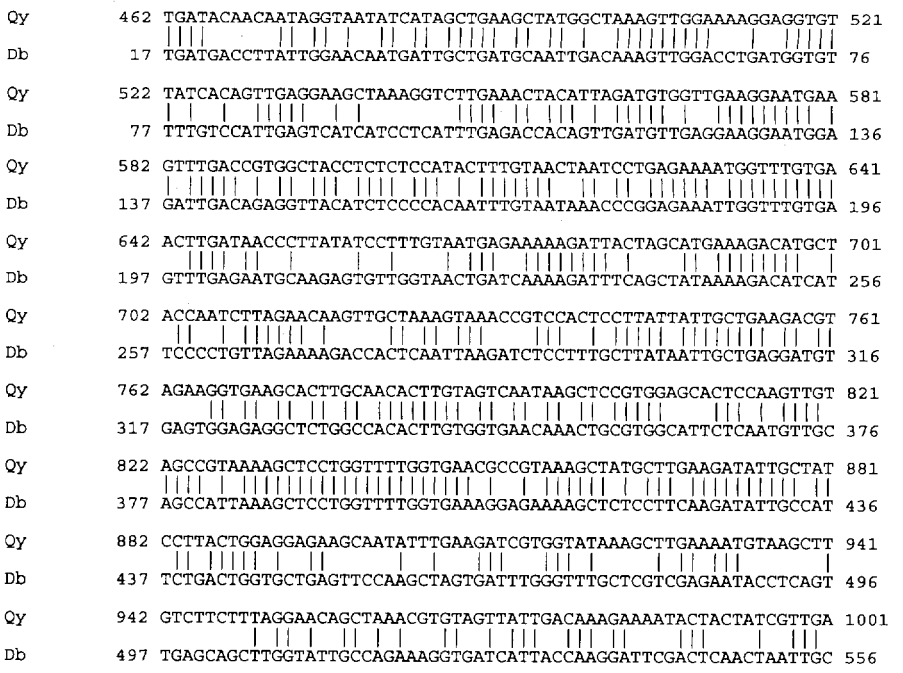
```
Qy          6 TTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAA 65
Db        164 TGCCAAAGATGTAAAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 223

Qy         66 ACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAA 125
Db        224 TTTAGCCGATGCTGTGGCCGTTACAATGGGGCCAAAGGGAAGAACAGTGATTATTGAGCA 283

Qy        126 GTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACT 185
Db        284 GAGTTGGGGAAGTCCCAAAGTAACAAAAGATGGTGTGACTKTTGCAAAGTCAATTGACTT 343

Qy        186 TGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAG 245
Db        344 AAAAGATAAATACAAAAACATTGGAGCTAAACTTGTTCAAGATGTTGCCAATAACACAAA 403

Qy        246 CGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGA 305
Db        404 TGAAGAAGCTGGGGATGGCACTACCACTGCTACTGTACTGGCACGCTCTATAGCCAAGGA 463

Qy        306 AGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAA 365
Db        464 AGGCTTCGAGAAGATTAGCAAAGGTGCTAATCCAGTGGAAATCAGGAGAGGTGTGATGTT 523

Qy        366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 425
Db        524 AGCTGTTGATGCTGTTATTGCTGAACTTAAAAAGCAGTTTAAACCTGTTACCACCCCTGA 583

Qy        426 AGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCAT 485
Db        584 AGAAATTGCACAGGTTGCTACGATTTCTGCAAACGGAGACAAAGAAATTGGCAATATCAT 643

Qy        486 AGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
Db        644 CTCTGATGCAATGAAAAAAGTTGGAAGAAAGGGTGTCATCACAGTTAAGGATGGAAAAAC 703

Qy        546 TCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
Db        704 ACTGAATGATGAATTAGAAATTATTGAAGGCATGAAGTTTGATCGAGGCTATATTTCTCC 763

Qy        606 ATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTG 665
Db        764 ATACTTTATTAATACATCAAAAGGTCAGAAATGTGAATTCCAGGATGCCTATGTTCTGTT 823

Qy        666 TAATGAGAAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAA 725
Db        824 GAGTGAAAAGAAAATTTCTAGTATCCAGTCCATTGTACCTGCTCTTGAGGKTGCCAATGC 883

Qy        726 AGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACT 785
Db        884 TCACCGTAAGCCTTTGGTCATAATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACT 943

Qy        786 TGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGG 845
Db        944 CGTCTTGAATAGGCTAAAGGTTGGTCTTCAGGTTGTGGCAGTKAAGGCTCCAGGGTTTGG 1003

Qy        846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATT 905
Db       1004 TGAM--ATAGAAAGAACARCTTAAAGATATGGCTATTGCTACTGGTGGTGCAGTGTTTTG 1061

Qy        906 TGAAGA 911
Db       1062 RVAAGA 1067
```

RESULT 13
AL518632
LOCUS AL518632 1201 bp mRNA linear EST 12-MAY-2003
DEFINITION AL518632 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone CS0DA009YI15 5-PRIME, mRNA sequence.
ACCESSION AL518632
VERSION AL518632.2 GI:30537130
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization

```
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12782125.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1228.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DA009AE08QP1&cluster=1228.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DA009AE08QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA009YI15"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      364 a    222 c    301 g    295 t     19 others
ORIGIN

  Query Match           19.1%;  Score 314.8;  DB 9;  Length 1201;
  Best Local Similarity   58.4%;  Pred. No. 4.7e-63;
  Matches  541;  Conservative    4;  Mismatches  381;  Indels    0;  Gaps    0;

Qy          6 TTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAA 65
              | | |||||  |    |||| |||  | |||||  |    |  |  | |||||||||
Db        189 TGCCAAAGATGTAAAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTGTAGACCT 248

Qy         66 ACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAA 125
                | ||  |||||||    || ||| | || || || ||  | |  ||  |||||||  |
Db        249 TTTAGCCGATGCTGTGGCCGTTACAATGGGGCCAAAGGGAAGAACAGTGATTATTGAGCA 308

Qy        126 GTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACT 185
              |  ||  ||    ||      | ||||||||||||||  ||||||||||   ||||||  |
Db        309 GAGTTGGGGAAGTCCCAAAGTAACAAAAGATGGTGTGACTGTTGCAAAGTCAATTGACTT 368

Qy        186 TGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAG 245
                |||||||  |   |||| || || ||| || | ||| |||| || ||    || || |
Db        369 AAAAGATAAATACAAAAACATTGGAGCTAAACTTGTTCAAGATGTTGCCAATAACACAAA 428

Qy        246 CGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGA 305
               ||    ||||| ||||| |||||  || || || || || ||||   ||||       ||
Db        429 TGAAGAAGCTGGGGATGGCACTACCACTGCTACTGTACTGGCACGCTCTATAGCCAAGGA 488

Qy        306 AGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAA 365
              |||  |  |     |         |||  ||||||  |||  ||  |   | ||  |
Db        489 AGGCTTCGAGAAGATTAGCAAAGGTGCTAATCCAGTGGAAATCAGGAGAGGTGTGATGTT 548

Qy        366 AGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAA 425
              |||||||  ||||||  |  |   |||||  |     |  || |||        ||   |
Db        549 AGCTGTTGATGCTGTAATTGCTGAACTTAAAAAGCAGTCTAAACCTGTGACCACCCCTGA 608

Qy        426 AGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCAT 485
              ||||||  ||  ||  ||||  ||  ||||||||||||   ||  |  |  |||  ||  ||||||||
Db        609 AGAAATTGCACAGGTTGCTACGATTTCTGCAAACGGAGACAAAGAAATTGGCAATATCAT 668

Qy        486 AGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGG 545
                ||||  ||  |||   ||||||||||  |   |||||  ||||||||   ||||  |  |||
Db        669 CTCTGATGCAATGAAAAAAGTTGGAAGAAAGGGTGTCATCACAGTAAAGGATGGAAAAAC 728

Qy        546 TCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCC 605
              ||  |    |  ||||||  |  |||||||  ||:  ||||||  || |||||  | |||||
Db        729 ACTGAATGATGAATTAGAAATTATTGAAGGCATSCCGTTTGATCGAGGCTATATTTCTCC 788

Qy        606 ATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTG 665
              |||||||  |  |  ||    |  |       |   ||||||  |  |  |  | |||  | ||  |
Db        789 ATACTTTATTAATACATCAAAAGGTCAGAAATGTGAATTCCAGGATGCCTATGTTCTGTT 848

Qy        666 TAATGAGAAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAA 725
              | |||  ||  ||  |||  ||||  ||  |   |||  ||||      | |||   |    |
Db        849 GAGTGAAAAGAAAATTTCTAGTATCCAGTCCATTGTACCTGCTCTTGAAATTGCCAATGC 908

Qy        726 AGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACT 785
                        ||   |   | ||  ||  ||||||||  ||  ||  ||  |||||  ||    |||||
Db        909 TCACCGTAAGCCTTTGGTCATAATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACT 968

Qy        786 TGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGG 845
              ||   | ||||  |||     |   |   ||  ||  |||||  ||  ||  ||  |||||  ||   ||  |
Db        969 CGTCTTGAATAGGCTAAAGGTTGGTCTTCAGGTTGTGGCAGTCAAGGCTCCAGGGGTTTG 1028

Qy        846 TGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATT 905
                ||     | ||     ||||  |||||||  |||||   ||||||  ||  |:      |
Db       1029 GTGACAATAGAYAGAMCAGCTTAAAGATATGGCTATTGCTACTGGTGGTGMGTGTTTGGA 1088

Qy        906 TGAAGATCGTGGTATAAAGCTTGAAA 931
                ||     ||    |  ||  ||::| |
Db       1089 GAGAGGGATTGACCTGAATCTKRAGA 1114


RESULT 14
BX439563
LOCUS       BX439563                1201 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX439563 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE012YC11
            5-PRIME, mRNA sequence.
ACCESSION   BX439563
VERSION     BX439563.1  GI:30773785
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1228.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DE012AB06QP1&cluster=1228.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DE012AB06QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE012YC11"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT      348 a    227 c    290 g    290 t     46 others
ORIGIN

  Query Match           19.0%;  Score 312.6;  DB 13;  Length 1201;
  Best Local Similarity   58.0%;  Pred. No. 1.6e-62;
  Matches  513;  Conservative   18;  Mismatches  353;  Indels    0;  Gaps    0;
```



```
RESULT 15
BF275584
LOCUS       BF275584                878 bp    mRNA    linear   EST 07-MAR-2001
DEFINITION  GA__Eb0024G23f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0024G23f, mRNA sequence.
ACCESSION   BF275584
VERSION     BF275584.1  GI:11206654
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0024G23f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      253 a    157 c    219 g    245 t      4 others
ORIGIN

  Query Match           19.0%;  Score 312.2;  DB 10;  Length 878;
  Best Local Similarity  61.8%;  Pred. No. 1.7e-62;
  Matches  494;  Conservative    0;  Mismatches  306;  Indels    0;  Gaps    0;
```



```
Qy       1002 TGGTGCTGGAAAATCAGAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGA 1061
Db        557 TGANGCAGCCTCAAAGGATGAGATACANACTAGGGTGGCACAACTAAAAAAAGAGCTAGC 616

Qy       1062 AGAAACAAGCTCAGATTATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGG 1121
Db        617 TCAGACAGATTCTGTCTATGATTCAGAAAAATTGGCAGAAAGGATTGCCAAACTATCTGG 676

Qy       1122 TGGAGTAGCTGTTATCCATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGA 1181
Db        677 TGGTGTTGCAGTCATTAAGGTGGGGGCTGCAACAGAGACTGAACTTGAGGATCGTAAGCT 736

Qy       1182 TCGTGTAGAAGATGCTCTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGG 1241
Db        737 ACGGATTGAAGATGCAAAGAATGCTACATTTGCTGCCATAGNAGAAGGTATTGTGCCTGG 796

Qy       1242 TGGTGGTACTGCTTTTGTCC 1261
Db        797 TGGTGGTGCTGCCTTAGTTC 816
```

Search completed: January 29, 2004, 15:25:33
Job time : 2474 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 14:44:13 ; Search time 475 Seconds
(without alignments)
12637.020 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 1647
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

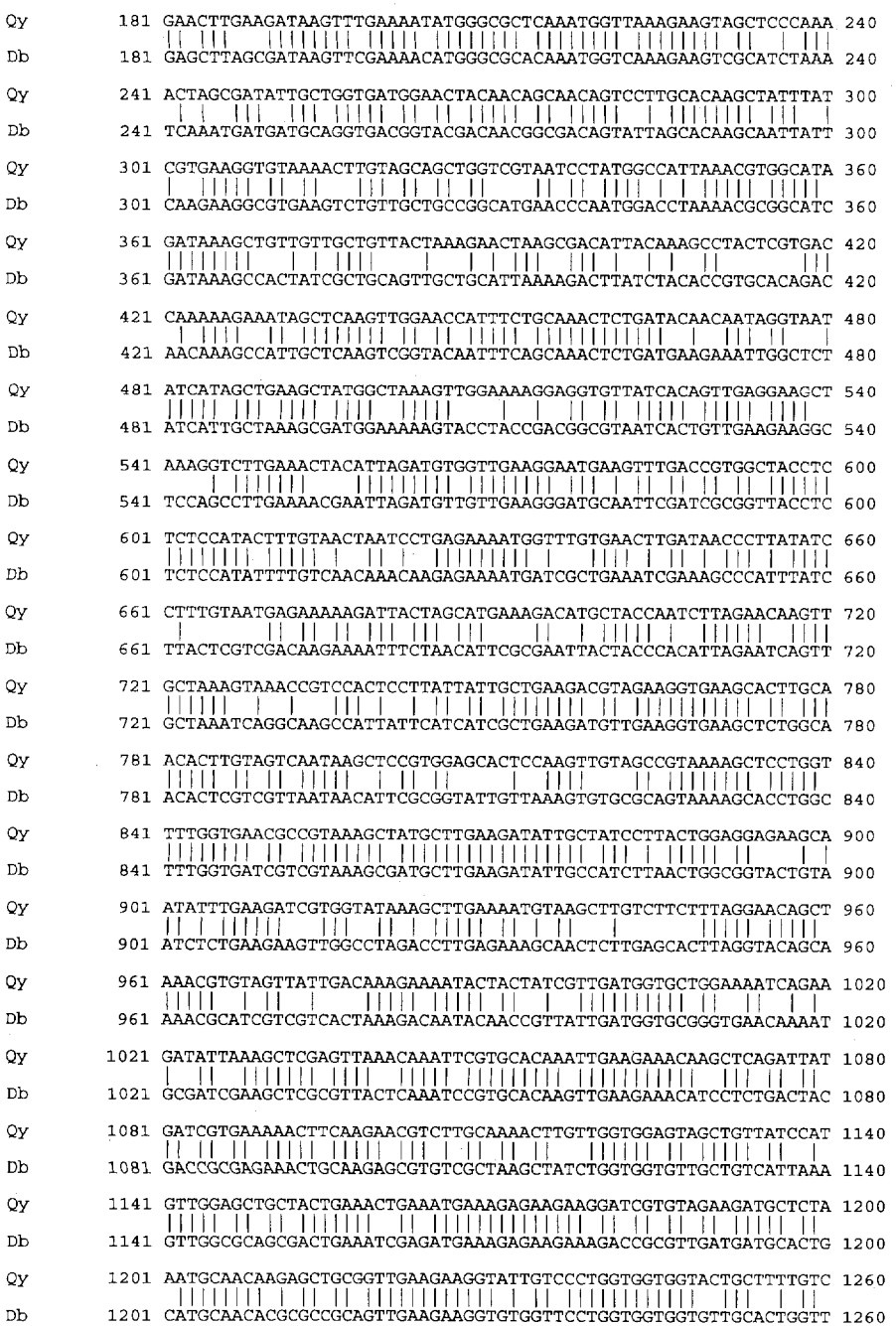
Result No.	Score	% Query Match	Length	DB	ID	Description
1	739.4	44.9	1635	13	US-10-228-167A-1	Sequence 1, Appli
2	725.8	44.1	1830121	15	US-10-329-960-1	Sequence 1, Appli
3	718.2	43.6	1626	12	US-10-369-493-33982	Sequence 33982, A
4	710	43.1	640681	10	US-09-790-988-1	Sequence 1, Appli
5	701.4	42.6	1635	12	US-10-369-493-40819	Sequence 40819, A
6	698.4	42.4	3625	10	US-09-070-927A-42	Sequence 42, Appl
c 7	690.2	41.9	1230025	12	US-10-289-762-1	Sequence 1, Appli
8	689.8	41.9	1926	15	US-10-267-311-50	Sequence 50, Appl
9	689.6	41.9	1635	9	US-09-841-132-380	Sequence 380, App
10	684.4	41.6	2155	10	US-09-960-428-13	Sequence 13, Appl
11	684.4	41.6	3840	12	US-10-157-317-26	Sequence 26, Appl
12	684.4	41.6	3840	12	US-10-157-339-26	Sequence 26, Appl
13	684.4	41.6	3840	13	US-10-157-305A-26	Sequence 26, Appl
14	684.4	41.6	3840	13	US-10-157-391-26	Sequence 26, Appl
15	684.4	41.6	3840	13	US-10-157-096-26	Sequence 26, Appl
16	684.4	41.6	3840	13	US-10-157-215A-26	Sequence 26, Appl
17	684.4	41.6	3840	13	US-10-157-302-26	Sequence 26, Appl
18	684.4	41.6	3840	13	US-10-154-951B-26	Sequence 26, Appl
19	684.4	41.6	3840	13	US-10-157-299-26	Sequence 26, Appl
20	684.4	41.6	3840	13	US-10-156-831-26	Sequence 26, Appl
21	684.4	41.6	3840	13	US-10-156-902-26	Sequence 26, Appl
22	684.4	41.6	3840	13	US-10-157-147-26	Sequence 26, Appl
23	684.4	41.6	3840	13	US-10-157-166-26	Sequence 26, Appl
24	684.4	41.6	3840	13	US-10-157-318-26	Sequence 26, Appl
25	684.4	41.6	3840	13	US-10-156-811-26	Sequence 26, Appl
26	684.4	41.6	3840	13	US-10-157-320A-26	Sequence 26, Appl
27	684.4	41.6	3840	13	US-10-157-418A-26	Sequence 26, Appl
28	683.8	41.5	1635	12	US-10-369-493-46781	Sequence 46781, A
29	681	41.3	1641	12	US-10-369-493-33701	Sequence 33701, A
30	680.8	41.3	1620	12	US-10-369-493-23868	Sequence 23868, A
31	643.8	39.1	1644	12	US-10-369-493-41211	Sequence 41211, A
32	639.2	38.8	1724	13	US-09-882-227-417	Sequence 417, App
33	637.4	38.7	1623	12	US-10-369-493-32974	Sequence 32974, A
34	629.2	38.2	1736	13	US-09-841-260-19	Sequence 19, Appl
35	629.2	38.2	1736	14	US-10-007-693-19	Sequence 19, Appl
36	623.2	37.8	1638	12	US-10-369-493-23854	Sequence 23854, A
37	607.6	36.9	1632	12	US-10-369-493-35447	Sequence 35447, A
38	604.4	36.7	1629	12	US-10-369-493-38952	Sequence 38952, A
39	587.4	35.7	1653	12	US-10-369-493-43871	Sequence 43871, A
40	585.2	35.5	1647	12	US-10-369-493-34549	Sequence 34549, A
41	563.4	34.2	1653	12	US-10-369-493-42681	Sequence 42681, A
42	560	34.0	1632	12	US-10-369-493-39155	Sequence 39155, A
43	560	34.0	1632	12	US-10-369-493-39904	Sequence 39904, A
44	553.6	33.6	1704	13	US-10-032-585-6410	Sequence 6410, Ap
45	551.8	33.5	1620	12	US-10-369-493-39525	Sequence 39525, A

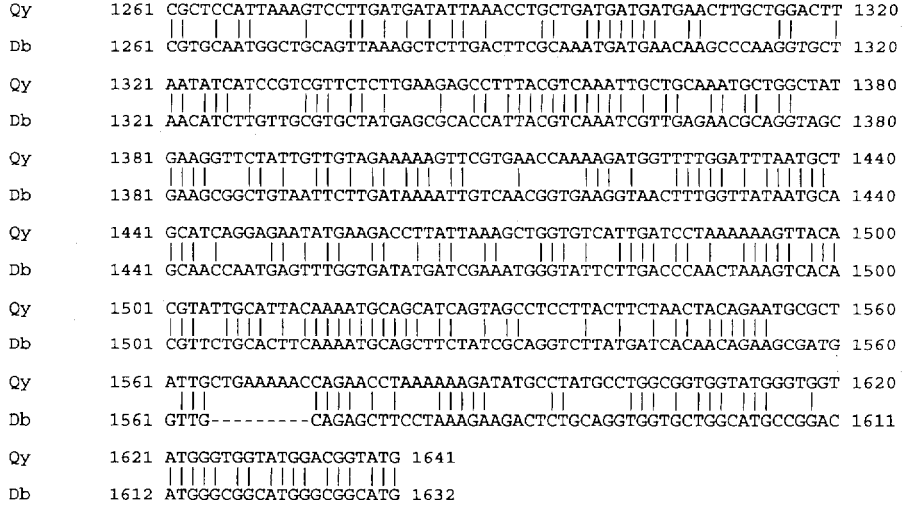
ALIGNMENTS

RESULT 1
US-10-228-167A-1
; Sequence 1, Application US/10228167A
; Publication No. US20030147909A1
; GENERAL INFORMATION:
; APPLICANT: Hernan Marshall Gonzalez, Sergio
; TITLE OF INVENTION: HIGHLY IMMUNOGENIC PROTEIN AGAINST THE INTRACELLULAR PATHOGEN AGEN
; TITLE OF INVENTION: PISCIRICKETTSIA SALMONIS, WHICH AFFECTS SALMON CULTURE, AMINO ACI
; TITLE OF INVENTION: ACID SEQUENCES OF SAID PROTEIN AND ITS APPLICATION IN THE DEVELOI
; TITLE OF INVENTION: METHODS FOR THE PREVENTION AND DIAGNOSIS OF DISEASES CAUSED BY SI
; FILE REFERENCE: 076502-9004
; CURRENT APPLICATION NUMBER: US/10/228,167A
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: Chilean Patent Application No. US20030147909A1 2086-2001
; PRIOR FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Piscirickettsia salmonis
US-10-228-167A-1

Query Match 44.9%; Score 739.4; DB 13; Length 1635;
Best Local Similarity 66.2%; Pred. No. 2.7e-141;
Matches 1086; Conservative 0; Mismatches 546; Indels 9; Gaps 1;

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              |||  |  |||||| | |  || | | |      |||||  |||||  | |      |||||
Db          1 ATGTCAGCAAAAGAAGTGCGCTTCGGTACCGGTTCCCGTCAAAAAATGTTGGACGGTGTT 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                |   ||||||||||||| || ||||| || || || ||     || || || ||  || |
Db         61 AACCTACTTGCAAATGCAGTGAAAGTCACTCTAGGTCCACGCGGACGCAACGTTATTTTA 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              |||||||| ||||||  |||||    || || ||||||||||||||||||||  ||||||||
Db        121 GAAAAGTCATTTGGTGCCCCAACCATCACTAAAGATGGTGTATCTGTTGCCAAAGAAATC 180
```





RESULT 2
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:

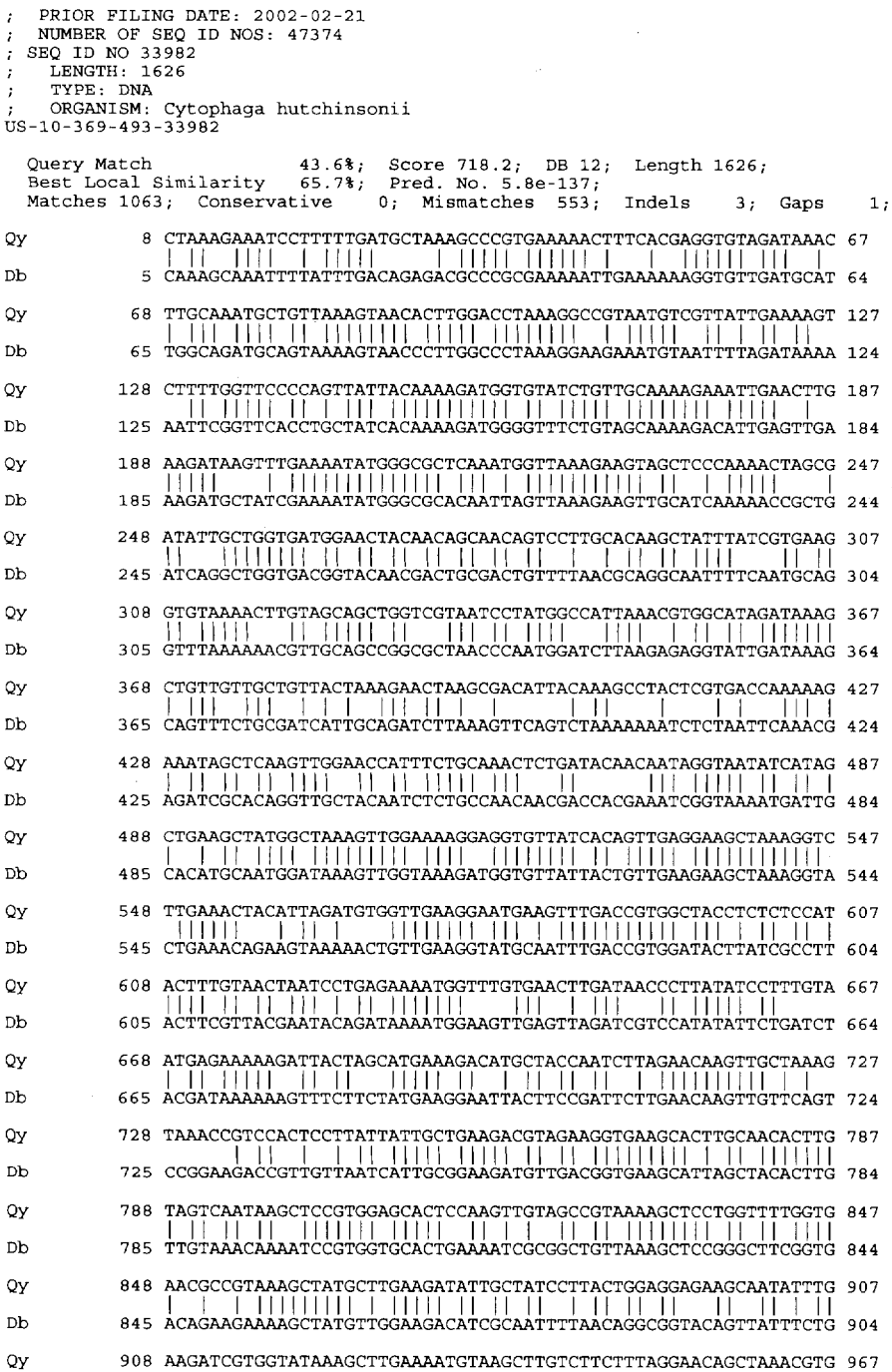
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)

```
;    OTHER INFORMATION: n equals a, t, g or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (150841)..(150841)
;    OTHER INFORMATION: n equals a, t, g or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (152500)..(152500)
;    OTHER INFORMATION: n equals a, t, g or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (152530)..(152530)

  Query Match             44.1%;  Score 725.8;  DB 15;  Length 1830121;
  Best Local Similarity   65.9%;  Pred. No. 2.7e-137;
  Matches 1086;  Conservative    0;  Mismatches  557;  Indels    6;  Gaps    2;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db     565350 ATGGCAGCAAAAGACGTAAAATTTGGTAATGACGCACGCGTAAAAAATGCTTAAAGGCGTG 565409

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db     565410 AATGTATTAGCGGATGCAGTAAAAGTAACCCTTGGCCCGAAAGGTCGTCATGTAATTTTA 565469

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db     565470 GATAAATCATTTGGCGCACCAACTATCACTAAAGACGGTGTGTCTGTTGCTCGTGAAATC 565529

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db     565530 GAATTAGAAGATAAATTCGAAAACATGGGGCACAAATGGTGAAAGAAGTGGCATCTAAA 565589

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db     565590 GCAAATGATGCTGCAGGTGACGGTACAACTACTGCAACTGTACTTGCGCAAGCTATCGTA 565649

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db     565650 AATGAAGGCTTGAAAGCAGTAGCTGCAGGTATGAATCCAATGGATTTAAAACGTGGTATT 565709

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db     565710 GATAAAGCAGTAAGTGCGGTTGTTTCTGAACTTAAAAATTTATCTAAACCTTGTGAAACA 565769

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db     565770 GCAAAAGAAATTGAACAAGTAGGGACTATTTCTGCAAACTCTGACAGCATTGTGGGTCAA 565829

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db     565830 TTAATCTCACAAGCAATGGAAAAAGTGGGCAAGGAAGGTGTAATTACTGTTGAAGATGGT 565889

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db     565890 ACGGGTCTTGAAGATGAATTAGATGTGGTTGAAGGCATGCAATTCGACCGTGGTTACCTT 565949

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db     565950 TCTCCATATTTCATCAACAAACCAGAAACTGCAACCGTTGAATTAGATAATCCATATCTT 566009

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db     566010 CTTCTTGTAGATAAAAAAATCTCTAACATTCGTGAATTACTTCCCGTGTTAGAAGGCGTT 566069

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db     566070 GCGAAAGCAGGTAAACCGTTATTAATCATCGCTGAAGACGTGGAAGGCGAAGCGCTTGCA 566129

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db     566130 ACCTTAGTGGTAAACACTATGCGCGGTATCGTGAAAGTTGCAGCCGTGAAAGCACCAGGT 566189

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db     566190 TTTGGTGATCGTCGTAAAGCGATGTTACAAGATATTGCAATTTTAACAGCGGGTACAGTG 566249

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db     566250 ATTTCTGAAGAAATTGGTATGGAGCTTGAAAAAGCAACATTGGAAGATTTAGGTCAAGCA 566309

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db     566310 AAACGTGTTGTTATCAATAAAGATAACACAACCATTATTGATGGTATCGGCGATGAAGCA 566369

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db     566370 CAAATCAAAGGTCGTGTGGCTCAAATTCGTCAGCAAATCGAAGAATCAACTTCTGACTAC 566429

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db     566430 GACAAAGAAAAACTCCAAGAACGCGTGGCTAAATTAGCAGGCGGTGTGGCAGTAATCAAA 566489

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db     566490 GTGGGCGCAGCAACTGAAGTTGAAATGAAAGAGAAAAAAGATCGTGTAGATGATGCATTA 566549

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db     566550 CATGCAACTCGTGCAGCGGTTGAAGAAGGTATCGTTGCTGGTGGTGGCGTTGCATTAGTT 566609

Qy       1261 CGCTC---CATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db     566610 CGCGCASSAGCGAAAGTTGCAGCAAGCCTAAAAGGTGACAACGAAGAACAAAATGTGGGT 566669

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db     566670 ATCAAACTTGCATTACGTGCTATGGAAGCGCCTTTACGTCAAATCGTCACTAACGCAGGT 566729

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437
Db     566730 GAAGAAGCTTCAGTTGTGGCAAGTGCGGTTAAAAATGGCGAAGGAAACTTTGGTTATAAT 566789

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db     566790 GCAGGAACAGAACAGTACGGCGATATGATCGAAATGGGTATCTTAGATCCAACTAAAGTC 566849

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db     566850 ACTCGTTCTGCATTGCAATTCGCAGCTTCTGTAGCAGGCTTAATGATTACCACTGAATGT 566909

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db     566910 ATGGTAACTGATCTTCCAAAAGATGATAAAG---CCGATTTAGGTGCTGCTGGAATGGGC 566966

Qy       1618 GGTATGGGTGGTATGGACGGTATGTACTA 1646
Db     566967 GGTATGGGTGGAATGGGCGGAATGATGTA 566995
```

RESULT 3
US-10-369-493-33982
; Sequence 33982, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33982
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Cytophaga hutchinsonii
US-10-369-493-33982

Query Match 43.6%; Score 718.2; DB 12; Length 1626;
Best Local Similarity 65.7%; Pred. No. 5.8e-137;
Matches 1063; Conservative 0; Mismatches 553; Indels 3; Gaps 1;



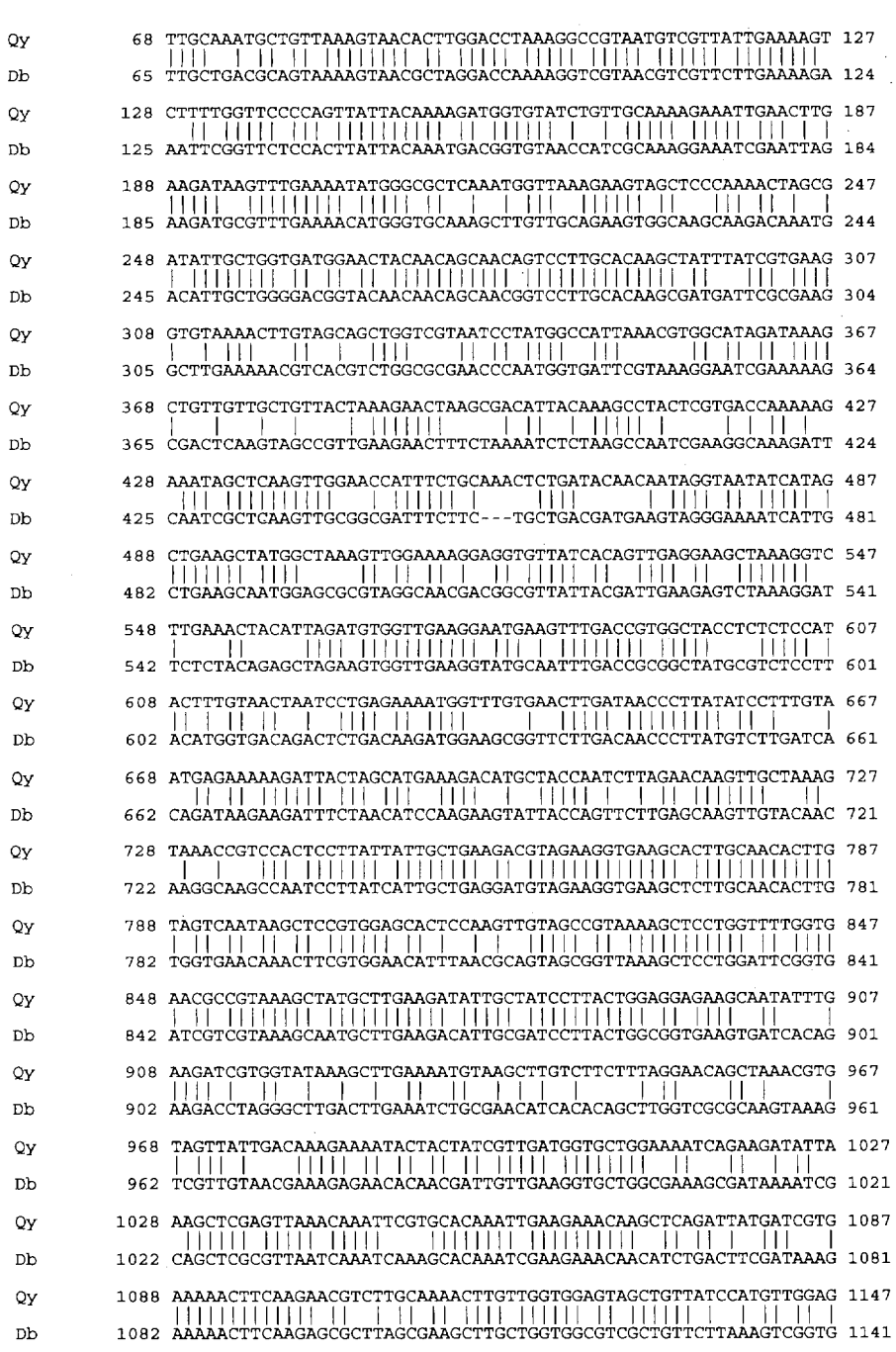
RESULT 4
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

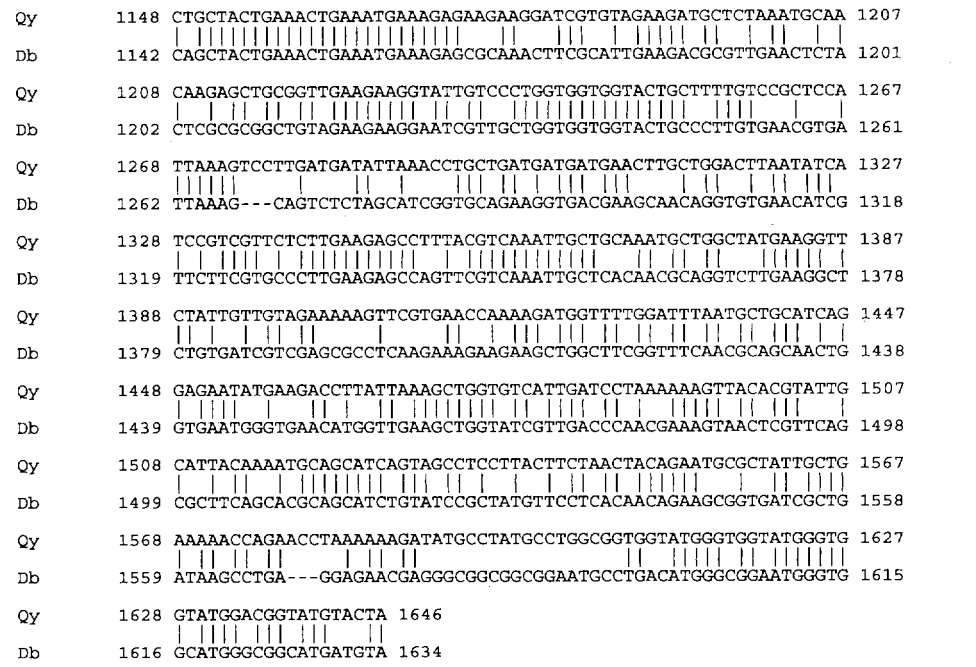
Query Match 43.1%; Score 710; DB 10; Length 640681;
Best Local Similarity 64.5%; Pred. No. 3.1e-134;

```
Matches 1061;  Conservative    0;  Mismatches  585;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db      18715 ATGGCCGCTAAAGATGTAAAATTTGGAAATGAAGCCCGCATTAAAATGCTTCGTGGAGTT 18774

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db      18775 AATGTATTAGCAGATGCAGTAAAAGTGACTTTAGGACCAAAAGGTAGAAATGTAGTTCTA 18834

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db      18835 GATAAATCTTTTGGAGCACCTAGTATTACTAAAGATGGTGTATCCGTAGCCCGTGAAATT 18894

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db      18895 GAATTAGAAGATAAATTCGAAAACATGGGAGCTCAAATGGTAAAAGAAGTTGCATCAAAA 18954

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db      18955 GCAAACGATGCAGCAGGTGATGGTACCACAACAGCAACATTATTAGCACAATCTATAGTA 19014

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db      19015 AATGAAGGTTTAAAAGCAGTAGCAGCTGGTATGAATCCAATGGATCTGAAACGTGGAATT 19074

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db      19075 GATAAAGCTGTTATCAGTGCTGTAGAAGAATTAAAACATTTATCTGTACCATGTTCTGAT 19134

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db      19135 TCTAAAGCAATTACACAAGTTGGTACTATTTCTGCAAATGCAGATGAAAAAGTTGGTTCT 19194

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db      19195 TTAATTGCAGAAGCAATGGAAAAAGTTGGTAATGACGGAGTTATTACAGTAGAAGAAGGT 19254

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db      19255 ACAGGTTTACAGGATGAACTTGAAGTTGTCAAAGGGATGCAATTTGATCGTGGTTATCTA 19314

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db      19315 TCTCCATATTTTATCAATAAACCAGAAACAGGTATTGTTGAATTAGAAAACCCATATATT 19374

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db      19375 TTAATGGCTGATAAAAAAATATCTAATGTTCGTGAAATGTTACCAATATTAGAATCTGTT 19434

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db      19435 GCAAAATCAGGAAAACCACTATTAATTATTTCTGAAGATTTAGAGGGTGAAGCTTTAGCA 19494

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db      19495 ACTTTAGTAGTTAATTCAATGAGAGGAATTGTAAAAGTCGCAGCAGTAAAAGCACCTGGA 19554

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db      19555 TTTGGTGATCGTCGTAAAGCAATGTTACAAGATATTTCAATTCTTACTGGTGGTTCTGTT 19614

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db      19615 ATCTCTGAAGAATTAGCTATGGAATTAGAAAAATCTACTTTAGAAGATTTAGGACAAGCA 19674

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db      19675 AAACGTGTTGTTATTAGCAAAGACACTACAACTATTATTGGTGGTGTAGGAGAAAAACAC 19734

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db      19735 TCCATTCAAAGTCGTATTAGTCAAATCCGACAAGAAATTCAAGAAGCTACTTCTGATTAT 19794

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db      19795 GATAAAGAAAAATTAAATGAACGCTTAGCTAAACTATCAGGCGGTGTTGCAGTACTAAAA 19854

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db      19855 GTAGGTGCGGCTACAGAAGTAGAAATGAAAGAAAAAAAAGCTCGTGTTGAAGATGCATTA 19914

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db      19915 CATGCTACTCGTGCAGCTGTAGAAGAAGGTGTAGTTGCTGGAGGTGGTGTTGCATTAGTG 19974

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db      19975 CGTGTAGCAGGGAAAATAGCTGATTTACGTGGTCAAAATGAAGATCAGAACGTAGGTATT 20034

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db      20035 CGAGTTGCTTTGCGTGCAATGGAAGCTCCATTACGTCAAATTGTTTCTAATTCTGGTGAA 20094

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db      20095 GAACCTTCTGTAGTTACAAACAATGTAAAAGACGGAAAAGGTAACTATGGTTACAATGCA 20154

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db      20155 GCTACTGATGAATATGGTGACATGATAGATTTTGGTATATTAGATCCAACTAAAGTTACA 20214

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db      20215 CGTTCTGCTTTACAGTATGCTGCTTCTGTCGCTGGTCTAATGATCACAACAGAATGTATG 20274

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db      20275 GTAACTGACTTGCCTAAAGAAGATAAATCTTCTGATTCTAGTTCTTCTCCAGCAGGAGGA 20334

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
Db      20335 ATGGGTGGCATGGGTGGAATGATGTA 20360
```

RESULT 5
US-10-369-493-40819
; Sequence 40819, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40819
;   LENGTH: 1635
;   TYPE: DNA
;   ORGANISM: Bacillus halodurans
US-10-369-493-40819

```
  Query Match             42.6%;  Score 701.4;  DB 12;  Length 1635;
  Best Local Similarity   65.5%;  Pred. No. 1.6e-133;
  Matches 1074;  Conservative    0;  Mismatches  556;  Indels    9;  Gaps    3;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db          5 CAAAAGATATTAAGTTTAGCGAAGACGCACGTCGCTCTATGCTTCGTGGTGTAGATAAAC 64
```





RESULT 6
US-09-070-927A-42
; Sequence 42, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302

```
;           REFERENCE/DOCKET NUMBER: PB369
;        TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (301) 309-8504
;           TELEFAX: (301) 309-8512
;     INFORMATION FOR SEQ ID NO: 42:
;        SEQUENCE CHARACTERISTICS:
;           LENGTH: 3625 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-070-927A-42

  Query Match             42.4%;  Score 698.4;  DB 10;  Length 3625;
  Best Local Similarity   66.1%;  Pred. No. 8.9e-133;
  Matches 1040;  Conservative    0;  Mismatches  527;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db        368 CAAAAGAGNTTAAATTTGCAGAAGATGCACGTGCAGCAATGCTACGCGGAGTAGATGTAT 427

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db        428 TAGCAGATACAGTGAAAGTGACATTAGGCCCTAAAGGTCGTAACGTTGTTTTAGAAAAAT 487

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        488 CATTTGGTTCACCATTGATTACTAACGATGGAGTAACAATTGCTAAGGAAATTGAATTGG 547

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        548 AAGATCATTTTGAAAACATGGGCGCAAAATTAGTTTCAGAAGTTGCTTCTAAAACGAATG 607

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        608 ATATTGCTGGTGACGGAACAACAACAGCGACTGTTTTGACACAAGCCATTGTTCGTGAAG 667

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        668 GCTTAAAAAACGTAACTGCTGGAGCCAACCCATTAGGTATTCGTCGTGGGATTGAATTAG 727

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        728 CAACAAAAACAGCAGTAGAAGAATTACACAATATTTCATCTGTAGTTGATTCAAAAGAAG 787

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        788 CGATTGCACAAGTCGCTGCTGTTTCATCAGGTTCTGA---AAAAGTCGGCCAATTAATTG 844

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        845 CCGATGCAATGGAAAAAGTTGGTAACGACGGCGTAATTACCATTGAAGAATCAAAAGGGA 904

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db        905 TTGAAACAGAATTAGATGTGGTTGAAGGAATGCAATTCGACCGCGGTTATTTATCTCAAT 964

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db        965 ACATGGTTACTGACAACGATAAAATGGAAGCTGTTTTAGAAAATCCATATATCTTAATTA 1024

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db       1025 CCGACAAAAAAATCTCAAATATTCAAGATATCTTACCTTTATTAGAACAAATTCTACAAC 1084

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db       1085 AAAGCCGTCCACTATTGATTATTGCGGATGATGTTGATGGGGAAGCTCTACCAACATTAG 1144

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db       1145 TATTGAACAAAATCCGTGGTACATTTAATGTTGTCGCAGTAAAAGCGCCAGGATTTGGTG 1204

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db       1205 ACCGCCGCAAAGCGATGCTTGAAGATATTGCTATTTTAACAGGTGGTACAGTAATCACTG 1264

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db       1265 ACGACTTAGGGTTAGAGTTAAAAGACACAACTATTGAAAACTTAGGAAATGCTAGCAAAG 1324

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db       1325 TAGTTGTCGACAAAGATAACACAACAATTGTCGAAGGTGCTGGTTCAAAAGAAGCCATTG 1384

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1385 ATGCCCGCGTTCATTTAATTAAAAACCAAATCGGCGAAACAACGTCTGATTTTGATCGTG 1444

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1445 AAAAATTACAAGAACGTTTAGCTAAATTAGCTGGTGGGGTTGCTGTCGTTAAAGTCGGTG 1504

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1505 CTGCAACTGAAACAGAATTAAAAGAATTAAAATTACGAATTGAAGATGCATTAAACGCAA 1564

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1565 CACGTGCCGCTGTAGAAGAAGGCATGGTTTCTGGTGGTGGTACCGCACTGGTCAATGTAA 1624

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1625 TTGGTAAAGTCGCTGCGCTAGAA---GCTGAAGGCGATGTGGCAACAGGGATCAAGATTG 1681

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1682 TCGTTCGTGCATTAGAAGAACCAATCCGTCAAATCGCTGAAAATGCTGGTTATGAAGGAT 1741

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1742 CAGTGATTGTTGACAAACTAAAAAATGTTGACTTAGGTATCGGATTCAATGCAGCTAACG 1801

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1802 GTGAATGGGTAAACATGGTTGAAGCCGGTATTGTTGACCCAACAAAAGTAACTCGTTCTG 1861

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       1862 CCTTACAAAATGCAGCTTCTGTGTCAGCTTTATTATTAACAACTGAAGCAGTTGTTGCAG 1921

Qy       1568 AAAAACCAGAACC 1580
Db       1922 ACAAACCAGAACC 1934

RESULT 7
US-10-289-762-1/c
; Sequence 1, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/10/289,762
;  CURRENT FILING DATE:  2003-03-27
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
;   LENGTH: 1230025
;   TYPE: DNA
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(15000)
```

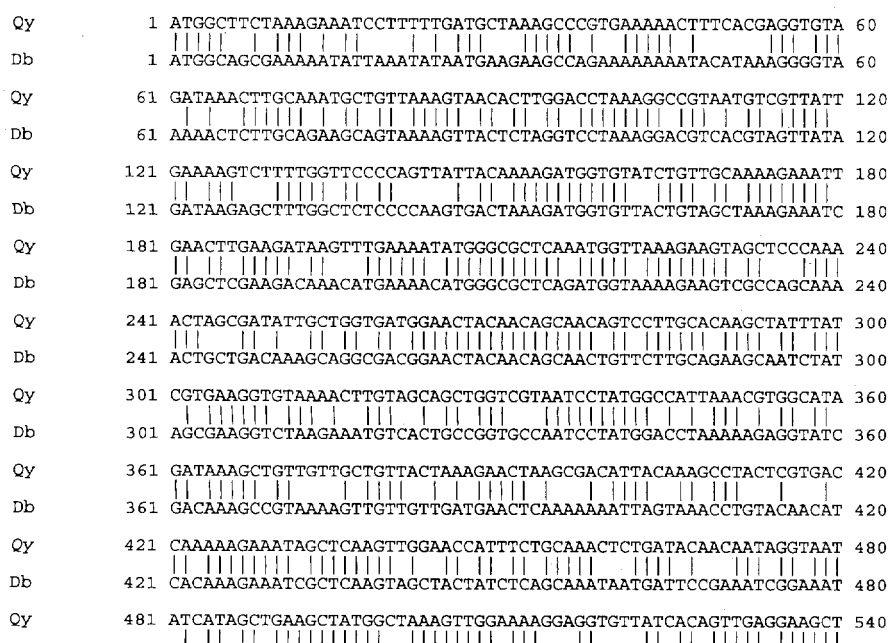
```
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15001)..(30000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (30001)..(45000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45001)..(60000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60001)..(75000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (75001)..(90000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90001)..(105000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (105001)..(120000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (120001)..(135000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (135001)..(150000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150001)..(165000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (165001)..(180000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (180001)..(195000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (195001)..(210000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (210001)..(225000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (225001)..(240000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (240001)..(255000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (255001)..(270000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (270001)..(285000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (285001)..(300000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (300001)..(315000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (315001)..(330000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (330001)..(345000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (345001)..(360000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (360001)..(375000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (375001)..(390000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (390001)..(405000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (405001)..(420000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (420001)..(435000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (435001)..(450000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (450001)..(465000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (465001)..(480000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (480001)..(495000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (495001)..(510000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (510001)..(525000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (525001)..(540000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (540001)..(555000)
;   OTHER INFORMATION: n=a or c or g or t
;   FEATURE:
```

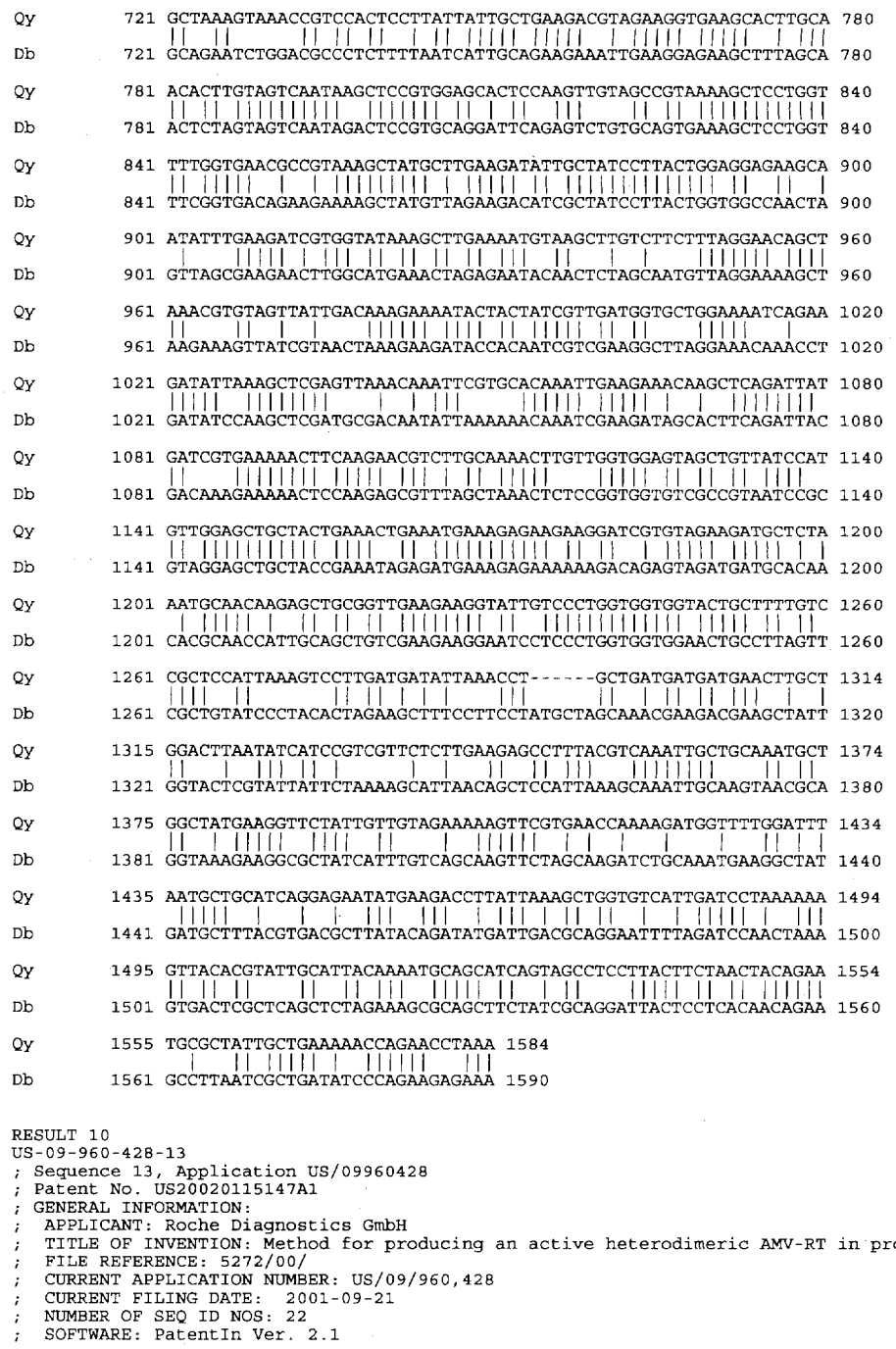
```
;    NAME/KEY: misc_feature
;    LOCATION: (555001)..(570000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (570001)..(585000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (585001)..(600000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (600001)..(615000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (615001)..(630000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (630001)..(645000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (645001)..(660000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (660001)..(675000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (675001)..(690000)
;    OTHER INFORMATION: n=a or c or g or t
;    FEATURE:

  Query Match             41.9%;  Score 690.2;  DB 12;  Length 1230025;
  Best Local Similarity   65.1%;  Pred. No. 4.4e-130;
  Matches 1035;  Conservative    0;  Mismatches  549;  Indels    6;  Gaps    1;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db     179472 ATGGCAGCGAAAAATATTAAATATAATGAAGAAGCCAGANAAAAAATACATAAAGGGGTA 179413

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db     179412 AAAACTCTTGCAGAAGCAGTAAAAGTTACTCTAGGTCCTAAAGGACGTCACGTAGTTATA 179353

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db     179352 GATAAGAGCTTTGGCTCTCCCCAAGTGACTAAAGATGGTGTTACTGTAGCTAAAGAAATC 179293

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db     179292 GAGCTCGAAGACAAACATGAAAACATGGGCGCTCAGATGGTAAAAGAAGTCGCCAGCAAA 179233

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db     179232 ACTGCTGACAAAGCAGGCGACGGAACTACAACAGCAACTGTTCTTGCAGAAGCAATCTAT 179173

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db     179172 AGCGAAGGTCTAAGAAATGTCACTGCCGGTGCCAATCCTATGGACCTAAAAAGAGGTATC 179113

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db     179112 GACAAAGCCGTAAAAGTTGTTGTTGATGAACTCAAAAAAATTAGTAAACCTGTACAACAT 179053

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db     179052 CACAAAGAAATCGCTCAAGTAGCTACTATCTCAGCAAATAATGATTCCGAAATCGGAAAT 178993

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db     178992 CTTATTGCAGAAGCTATGGAAAAAGTTGGTAAAAACGGATCCATTACTGTTGAAGAAGCT 178933

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db     178932 AAAGGCTTCGAAACTGTTCTCGACGTTGTAGAAGGAATGAACTTCAACCGTGGATACCTC 178873

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db     178872 TCCAGCTACTTCTCCACAAATCCAGAAACTCAAGAATGCGTTTTAGAAGACGCTCTGATT 178813

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db     178812 CTAATCTACGATAAAAAAATCTCTGGAATTAAAGACTTCCTTCCAGTTTTACAACAAGTA 178753

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db     178752 GCAGAATCTGGACGCCCTCTTTTAATCATTGCAGAAGAAATTGAAGGAGAAGCTTTAGCA 178693

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db     178692 ACTCTAGTAGTCAATAGACTCCGTGCAGGATTCAGAGTCTGTGCAGTGAAAGCTCCTGGT 178633

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db     178632 TTCGGTGACAGAAGAAAAGCTATGTTAGAAGACATCGCTATCCTTACTGGTGGCCAACTA 178573

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db     178572 GTTAGCGAAGAACTTGGCATGAAACTAGAGAATACAACTCTAGCAATGTTAGGAAAAGCT 178513

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db     178512 AAGAAAGTTATCGTAACTAAAGAAGATACCACAATCGTCGAAGGCTTAGGAAACAAACCT 178453

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db     178452 GATATCCAAGCTCGATGCGACAATATTAAAAAACAAATCGAAGATAGCACTTCAGATTAC 178393

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db     178392 GACAAAGAAAAACTCCAAGAGCGTTTAGCTAAACTCTCCGGTGGTGTCGCCGTAATCCGC 178333

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db     178332 GTAGGAGCTGCTACCGAAATAGAGATGAAAGAGAAAAAAGACAGAGTAGATGATGCACAA 178273

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db     178272 CACGCAACCATTGCAGCTGTCGAAGAAGGAATCCTCCCTGGTGGTGGAACTGCCTTAGTT 178213

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCT------GCTGATGATGATGAACTTGCT 1314
Db     178212 CGCTGTATCCCTACACTAGAAGCTTTCCTTCCTATGCTAGCAAACGAAGACGAAGCTATT 178153

Qy       1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCT 1374
Db     178152 GGTACTCGTATTATTCTAAAAGCATTAACAGCTCCATTAAAGCAAATTGCAAGTAACGCA 178093

Qy       1375 GGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTT 1434
Db     178092 GGTAAAGAAGGCGCTATCATTTGTCAGCAAGTTCTAGCAAGATCTGCAAATGAAGGCTAT 178033

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db     178032 GATGCTTTACGTGACGCTTATACAGATATGATTGACGCAGGAATTTTAGATCCAACTAAA 177973

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db     177972 GTGACTCGCTCAGCTCTAGAAAGCGCAGCTTCTATCGCAGGATTACTCCTCACAACAGAA 177913

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAA 1584
```

```
Db     177912 GCCTTAATCGCTGATATCCCAGAAGAGAAA 177883

RESULT 8
US-10-267-311-50
; Sequence 50, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
;  APPLICANT: Siegel, Marvin
;  APPLICANT:  Chu, N. Randall
;  APPLICANT:  Mizzen, Lee A.
;  TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
;  FILE REFERENCE: 12071/002001
;  CURRENT APPLICATION NUMBER: US/10/267,311
;  CURRENT FILING DATE:  2002-10-09
;  PRIOR APPLICATION NUMBER: US/09/613,303
;  PRIOR FILING DATE: 2000-07-10
;  PRIOR APPLICATION NUMBER: US 60/143,757
;  PRIOR FILING DATE: 1999-07-08
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: fusion sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1923)
US-10-267-311-50

  Query Match             41.9%;  Score 689.8;  DB 15;  Length 1926;
  Best Local Similarity   65.8%;  Pred. No. 4e-131;
  Matches 1035;  Conservative    0;  Mismatches  532;  Indels    6;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db          5 CAAAAGAAATTAAATTTTCATCAGATGCCCGTTCAGCTATGGTCCGTGGTGTCGATATCC 64

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db         65 TTGCAGATACTGTTAAAGTAACTTTGGGACCAAAAGGTCGCAATGTCGTTCTTGAAAAGT 124

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        125 CATTCGGTTCACCCTTGATTACCAATGACGGTGTGACTATTGCCAAAGAAATTGAATTAG 184

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        185 AAGACCATTTTGAAAATATGGGTGCCAAATTGGTATCAGAAGTAGCTTCAAAAACCAATG 244

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        245 ATATCGCAGGTGATGGAACTACAACTGCAACTGTTTTGACCCAAGCAATCGTCCGTGAAG 304

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        305 GAATCAAAAACGTCACAGCAGGTGCAAATCCAATCGGTATTCGTCGTGGGATTGAAACAG 364

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        365 CAGTTGCCGCAGCAGTTGAAGCTTTGAAAAACAACGTCATCCCTGTTGCCAATAAAGAAG 424

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        425 CTATCGCTCAAGTTGCAGCCGTATCTTCTCGTTCTGA---AAAAGTTGGTGAGTACATCT 481

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        482 CTGAAGCAATGGAAAAAGTTGGCAAAGACGGTGTCATCACCATCGAAGAGTCACGTGGTA 541

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
Db        542 TGGAAACAGAGCTTGAAGTCGTAGAAGGAATGCAGTTTGACCGTGGTTACCTTTCACAGT 601

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db        602 ACATGGTGACAGATAGCGAAAAAATGGTGGCTGACCTTGAAAATCCGTACATTTTGATTA 661

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db        662 CAGACAAGAAAATTTCCAATATCCAAGAAATCTTGCCACTTTTGGAAAGCATTCTCCAAA 721

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db        722 GCAATCGTCCACTCTTGATTATTGCGGATGATGTGGATGGTGAGGCTCTTCCAACTCTTG 781

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db        782 TTTTGAACAAGATTCGTGGAACCTTCAACGTAGTAGCAGTCAAGGCACCTGGTTTTGGTG 841

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db        842 ACCGTCGCAAAGCCATGCTTGAAGATATCGCCATCTTAACAGGCGGAACAGTTATCACAG 901

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db        902 AAGACCTTGGTCTTGAGTTGAAAGATGCGACAATTGAAGCTCTTGGTCAAGCAGCGAGAG 961

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db        962 TGACCGTGGACAAAGATAGCACGGTTATTGTAGAAGGTGCAGGAAATCCTGAAGCGATTT 1021

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1022 CTCACCGTGTTGCGGTTATCAAGTCTCAAATCGAAACTACAACTTCTGAATTTGACCGTG 1081

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1082 AAAAATTGCAAGAACGCTTGGCCAAATTGTCAGGTGGTGTAGCGGTTATTAAGGTCGGAG 1141

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1142 CCGCAACTGAAACTGAGTTGAAAGAAATGAAACTCCGCATTGAAGATGCCCTCAACGCTA 1201

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1202 CTCGTGCAGCTGTTGAAGAAGGTATTGTTGCAGGTGGTGGAACAGCTCTTGCCAATGTGA 1261

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1262 TTCCAGCTGTTGCTACCTTGGAATTGACAGGAGATGAAGCA---ACAGGACGTAATATTG 1318

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1319 TTCTCCGTGCTTTGGAAGAACCTGTTCGTCAAATTGCTCACAATGCAGGATTTGAAGGAT 1378

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1379 CTATCGTTATCGATCGTTTGAAAAATGCTGAGCTTGGTATAGGATTCAACGCAGCAACTG 1438

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1439 GCGAGTGGGTTAACATGATTGATCAAGGTATCATTGATCCAGTTAAAGTGAGTCGTTCAG 1498

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       1499 CCCTACAAAATGCAGCATCTGTAGCCAGCTTGATTTTGACAACAGAAGCAGTCGTAGCCA 1558

Qy       1568 AAAAACCAGAACC 1580
Db       1559 ATAAACCAGAACC 1571
```

RESULT 9
US-09-841-132-380
; Sequence 380, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 380
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
US-09-841-132-380

Query Match 41.9%; Score 689.6; DB 9; Length 1635;
Best Local Similarity 65.1%; Pred. No. 4.1e-131;
Matches 1035; Conservative 0; Mismatches 549; Indels 6; Gaps 1;

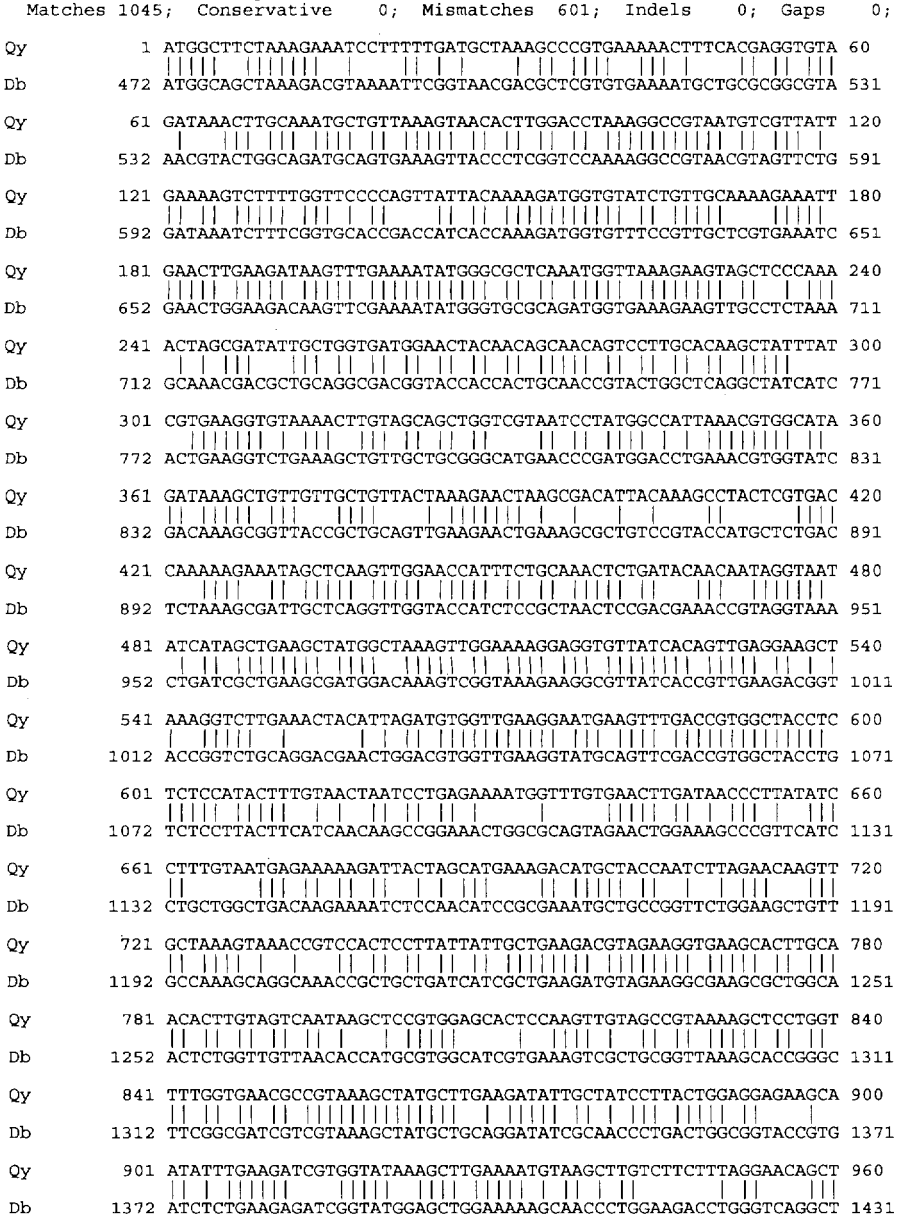


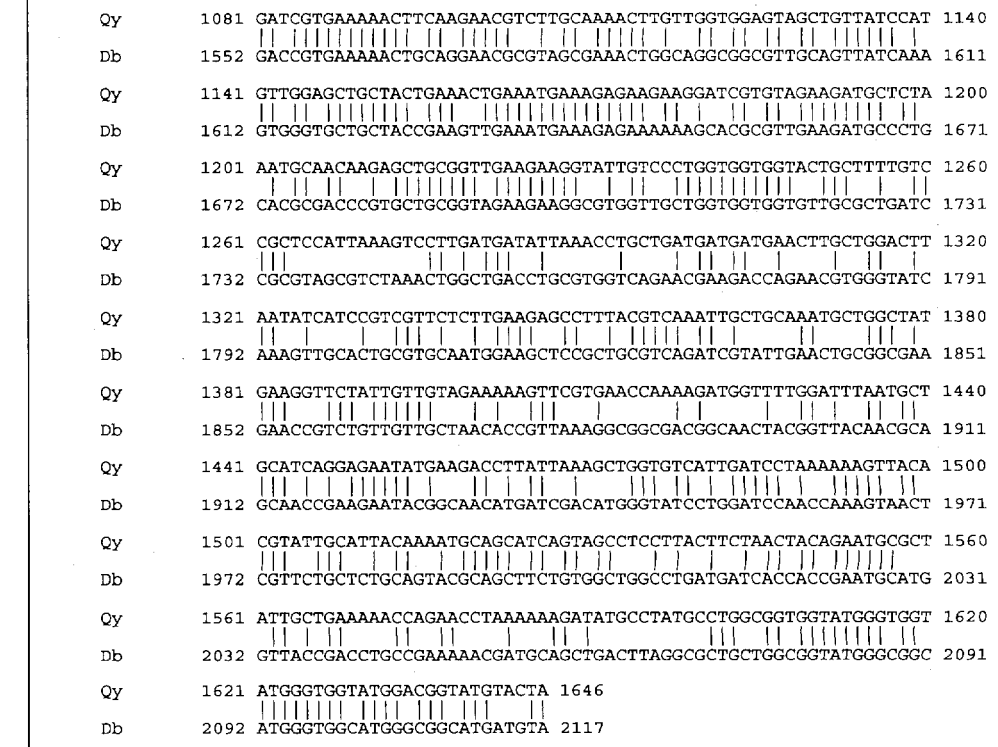


RESULT 10
US-09-960-428-13
; Sequence 13, Application US/09960428
; Patent No. US20020115147A1
; GENERAL INFORMATION:
; APPLICANT: Roche Diagnostics GmbH
; TITLE OF INVENTION: Method for producing an active heterodimeric AMV-RT in prokaryoti
; FILE REFERENCE: 5272/00/
; CURRENT APPLICATION NUMBER: US/09/960,428
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 13
; LENGTH: 2155
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-960-428-13

Query Match 41.6%; Score 684.4; DB 10; Length 2155;
Best Local Similarity 63.5%; Pred. No. 5.3e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;





RESULT 11
US-10-157-317-26
; Sequence 26, Application US/10157317
; Publication No. US20030232335A1
; GENERAL INFORMATION:
; APPLICANT: Surber, Mark W.
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Neil Berkley
; TITLE OF INVENTION: MINICELL-BASED SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: AND PROTEINS THAT MODULATE THE ACTIVITY OF SIGNALLING
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: MPEX.008DV21
; CURRENT APPLICATION NUMBER: US/10/157,317
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26

```
;   LENGTH: 3840
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Fusion protein
US-10-157-317-26

  Query Match             41.6%;  Score 684.4;  DB 12;  Length 3840;
  Best Local Similarity   63.5%;  Pred. No. 6.6e-130;
  Matches 1045;  Conservative    0;  Mismatches  601;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db       2133 ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA 2192

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db       2193 AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG 2252

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db       2253 GATAAATCTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC 2312

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db       2313 GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA 2372

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db       2373 GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC 2432

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       2433 ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC 2492

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       2493 GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC 2552

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       2553 TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 2612

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       2613 CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT 2672

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       2673 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG 2732

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       2733 TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC 2792

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       2793 CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT 2852

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       2853 GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA 2912

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       2913 ACTCTGGTTGTTAACACCATGCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC 2972

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       2973 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG 3032

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       3033 ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT 3092

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       3093 AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT 3152

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       3153 GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC 3212

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       3213 GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA 3272

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       3273 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG 3332

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       3333 CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC 3392

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       3393 CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC 3452

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       3453 AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA 3512

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       3513 GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA 3572

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       3573 GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT 3632

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       3633 CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG 3692

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db       3693 GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC 3752

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
Db       3753 ATGGGTGGCATGGGCGGCATGATGTA 3778
```

```
RESULT 12
US-10-157-339-26
; Sequence 26, Application US/10157339
; Publication No. US20040005700A1
; GENERAL INFORMATION:
;  APPLICANT: Surber, Mark W.
;  APPLICANT:  Matthew Giacalone
;  TITLE OF INVENTION: POROPLASTS
;  FILE REFERENCE: MPEX.008DV3
;  CURRENT APPLICATION NUMBER: US/10/157,339
;  CURRENT FILING DATE:  2002-08-29
;  PRIOR APPLICATION NUMBER: 60/293,566
;  PRIOR FILING DATE: 2001-05-24
;  PRIOR APPLICATION NUMBER: 60/359,843
;  PRIOR FILING DATE: 2002-02-25
;  PRIOR APPLICATION NUMBER: 10/154,951
;  PRIOR FILING DATE: 2002-05-24
;  NUMBER OF SEQ ID NOS: 257
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 3840
;   TYPE: DNA
```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion protein
US-10-157-339-26

Query Match 41.6%; Score 684.4; DB 12; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db       2133 ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA 2192

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db       2193 AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG 2252

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db       2253 GATAAATCTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC 2312

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db       2313 GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA 2372

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db       2373 GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC 2432

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db       2433 ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC 2492

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db       2493 GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC 2552

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db       2553 TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 2612

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db       2613 CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT 2672

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db       2673 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG 2732

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db       2733 TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC 2792

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       2793 CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT 2852

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       2853 GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA 2912

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       2913 ACTCTGGTTGTTAACACCATGCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC 2972

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       2973 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG 3032

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       3033 ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT 3092
```

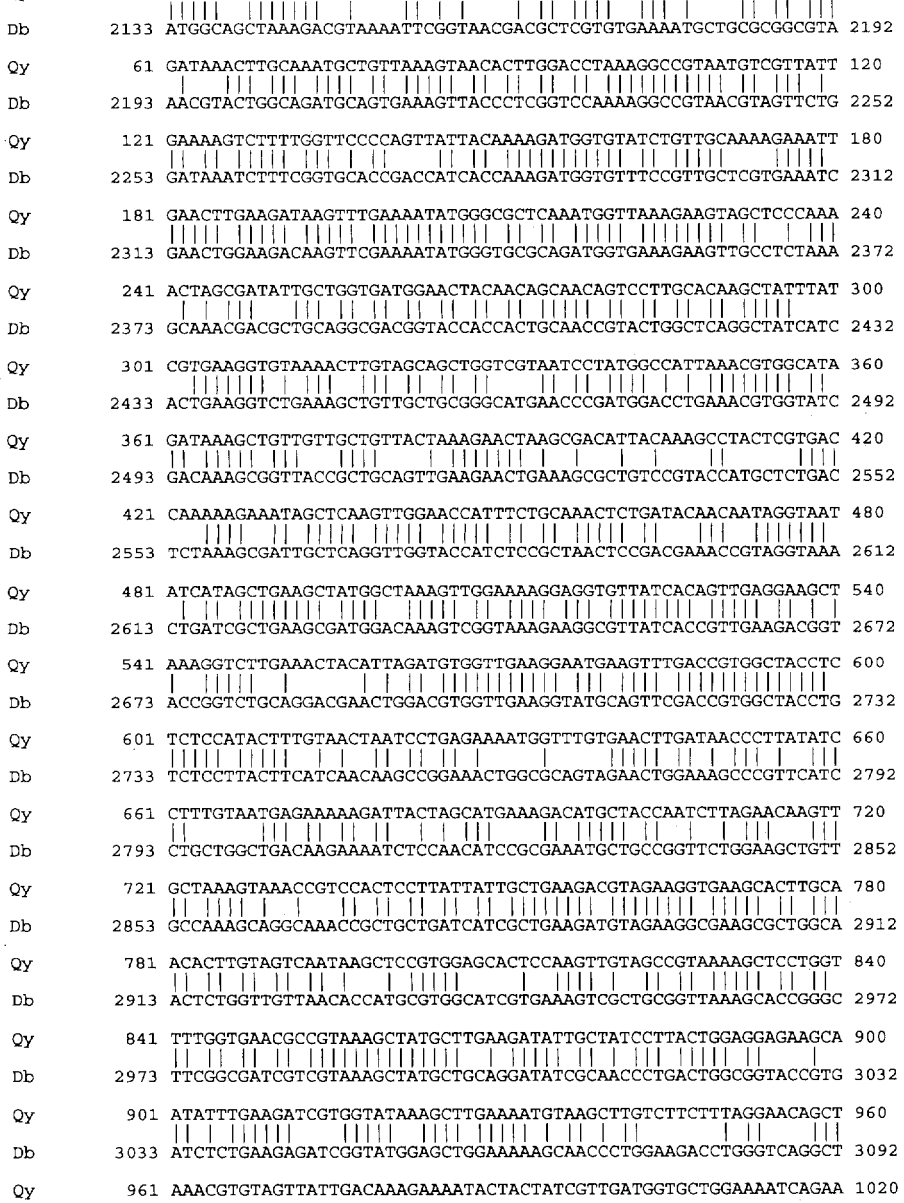
```
Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       3093 AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT 3152

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       3153 GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC 3212

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       3213 GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA 3272

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       3273 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG 3332

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       3333 CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC 3392

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       3393 CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC 3452

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       3453 AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA 3512

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       3513 GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA 3572

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       3573 GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT 3632

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       3633 CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG 3692

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db       3693 GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC 3752

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
Db       3753 ATGGGTGGCATGGGCGGCATGATGTA 3778
```

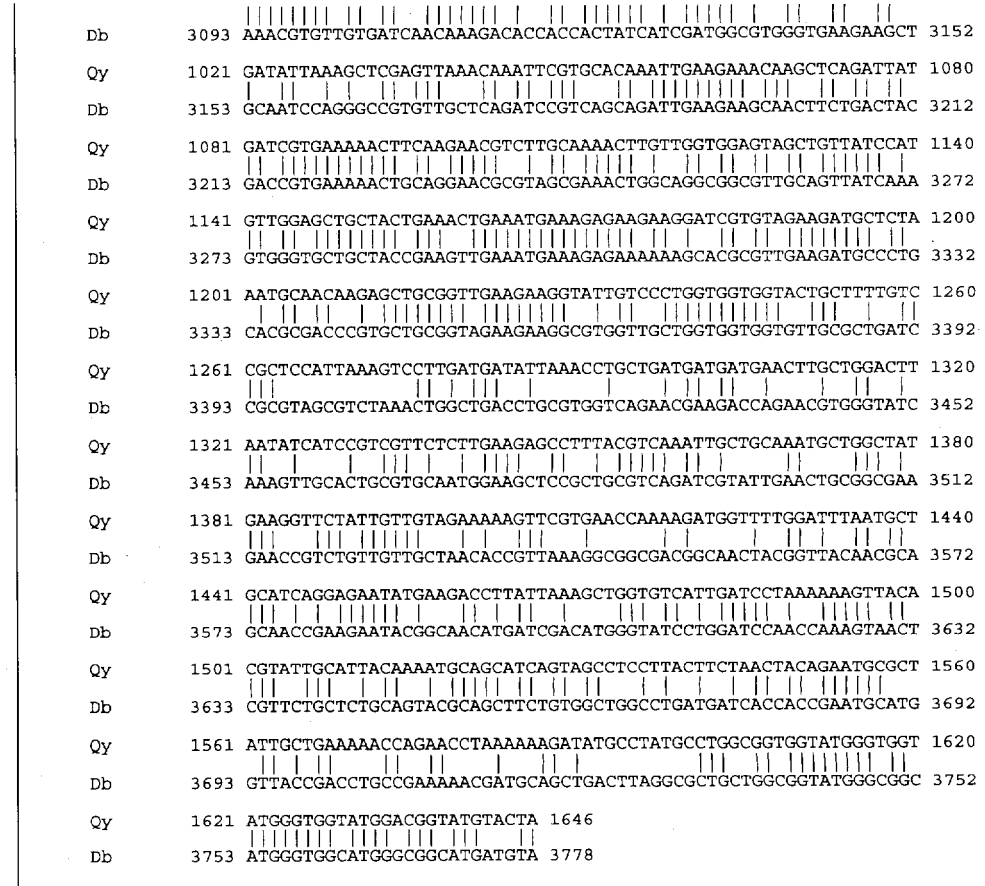
RESULT 13
US-10-157-305A-26
; Sequence 26, Application US/10157305A
; Publication No. US20030166099A1
; GENERAL INFORMATION:
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Surber, Mark W.
; APPLICANT: Neil Berkley
; APPLICANT: Anca M. Segall
; APPLICANT: Robert Klepper
; TITLE OF INVENTION: MINICELL COMPRISING MEMBRANE PROTEINS
; FILE REFERENCE: MPEX.008DV1
; CURRENT APPLICATION NUMBER: US/10/157,305A
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,951
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Gene encoding a fusion protein
US-10-157-305A-26

  Query Match           41.6%;  Score 684.4;  DB 13;  Length 3840;
  Best Local Similarity  63.5%;  Pred. No. 6.6e-130;
  Matches 1045;  Conservative    0;  Mismatches  601;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||| |||||||  |  || | |  | || ||||  ||| |  || || |||
Db       2133 ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA 2192

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                |  ||| ||| |||| || ||||| || || || || ||||||||||| || ||| |
Db       2193 AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG 2252

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              || || |||| ||| | ||    || || |||||||||||| || ||||    |||||
Db       2253 GATAAATCTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC 2312

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||| |||| |||| |||||||||||| || || |||| |||||||| ||  | |||
Db       2313 GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA 2372

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              | | |||   ||| || || || || || || || ||||| || || || || |||||
Db       2373 GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC 2432

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||  | |||  ||| || ||  || || |||| | |  |||||||| ||
Db       2433 ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC 2492

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              || ||||| |||   |||     | |||||| |  |   |   ||      ||||
Db       2493 GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC 2552

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                 ||||  || ||||| ||||| ||||| || || ||||| ||   |||  |||||||
Db       2553 TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 2612

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              | || |||||||| ||||  ||||| || |||| ||| |||||||| ||||| || | |
Db       2613 CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT 2672

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |  |||||  |       | | || |||||||||||| ||| |||| ||||||||||||||
Db       2673 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG 2732

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||| |||||  |  |  || || || |      |    |||||  || | |||  |  |||
Db       2733 TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC 2792

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||        ||| || || ||  | | |||     || ||||| ||  |   | |||   |||
Db       2793 CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT 2852

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              || |||| |  |   || || || || || |||||||| |||||||| ||||| || |||
Db       2853 GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA 2912

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              || || || || || |  |  ||||     | |||| |  || || ||||| || ||
Db       2913 ACTCTGGTTGTTAACACCATGCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC 2972

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              || || || ||||||||||||||  | ||||| || | ||| |||||  ||     |
Db       2973 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG 3032

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              || | ||||||    |||||  |||| ||||| | || | ||      |||     |||
Db       3033 ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT 3092

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||||||| || ||  |||||| | || |||||| | ||||| |  ||  ||  ||
Db       3093 AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT 3152

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              |  ||  | |  || |||   || || |||   || ||||||||| |||  || || ||
Db       3153 GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC 3212

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              || |||||||||| || |||||  | || |||||  |  || || || || |||||| |
Db       3213 GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA 3272

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              || || |||||||  |||  |||||||||||||||| || |  || || |||||||  ||
Db       3273 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG 3332

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              | || ||  | |||||||| ||||||||  | ||  |||||||||||  |||  |  ||
Db       3333 CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC 3392

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||              || | |||  |     |  | || ||  |   |  ||  |
Db       3393 CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC 3452

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||  |    |  ||| |  |||| ||  | ||||| || |    ||    ||| |
Db       3453 AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA 3512

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   ||| ||||||   | |  |||   |   | |    |  || |  || ||
Db       3513 GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA 3572

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||| | | |||||| |   || | ||  |    ||| || | ||||| |  ||||| ||
Db       3573 GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT 3632

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||  |||  | ||  | ||||| || || ||   | |  | | || || ||||||
Db       3633 CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG 3692

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              || | ||    ||  ||    | || |          |||  || |||||||| ||
Db       3693 GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC 3752

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
              |||||||| |||| ||| |||   ||
Db       3753 ATGGGTGGCATGGGCGGCATGATGTA 3778


RESULT 14
US-10-157-391-26
; Sequence 26, Application US/10157391
; Publication No. US20030166279A1
; GENERAL INFORMATION:
;  APPLICANT: Sabbadini, Roger A.
;  APPLICANT:  Neil Berkley
;  TITLE OF INVENTION: MINICELL-BASED TRANSFECTION
;  FILE REFERENCE: MPEX.008DV14
;  CURRENT APPLICATION NUMBER: US/10/157,391
;  CURRENT FILING DATE:  2002-05-28
;  PRIOR APPLICATION NUMBER: 60/293,566
;  PRIOR FILING DATE: 2001-05-24
;  PRIOR APPLICATION NUMBER: 60/359,843
;  PRIOR FILING DATE: 2002-02-25
;  PRIOR APPLICATION NUMBER: 10/154,951
;  PRIOR FILING DATE: 2002-05-24
;  NUMBER OF SEQ ID NOS: 257
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
;   LENGTH: 3840
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
```

; FEATURE:
; OTHER INFORMATION: Fusion protein
US-10-157-391-26

Query Match 41.6%; Score 684.4; DB 13; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;

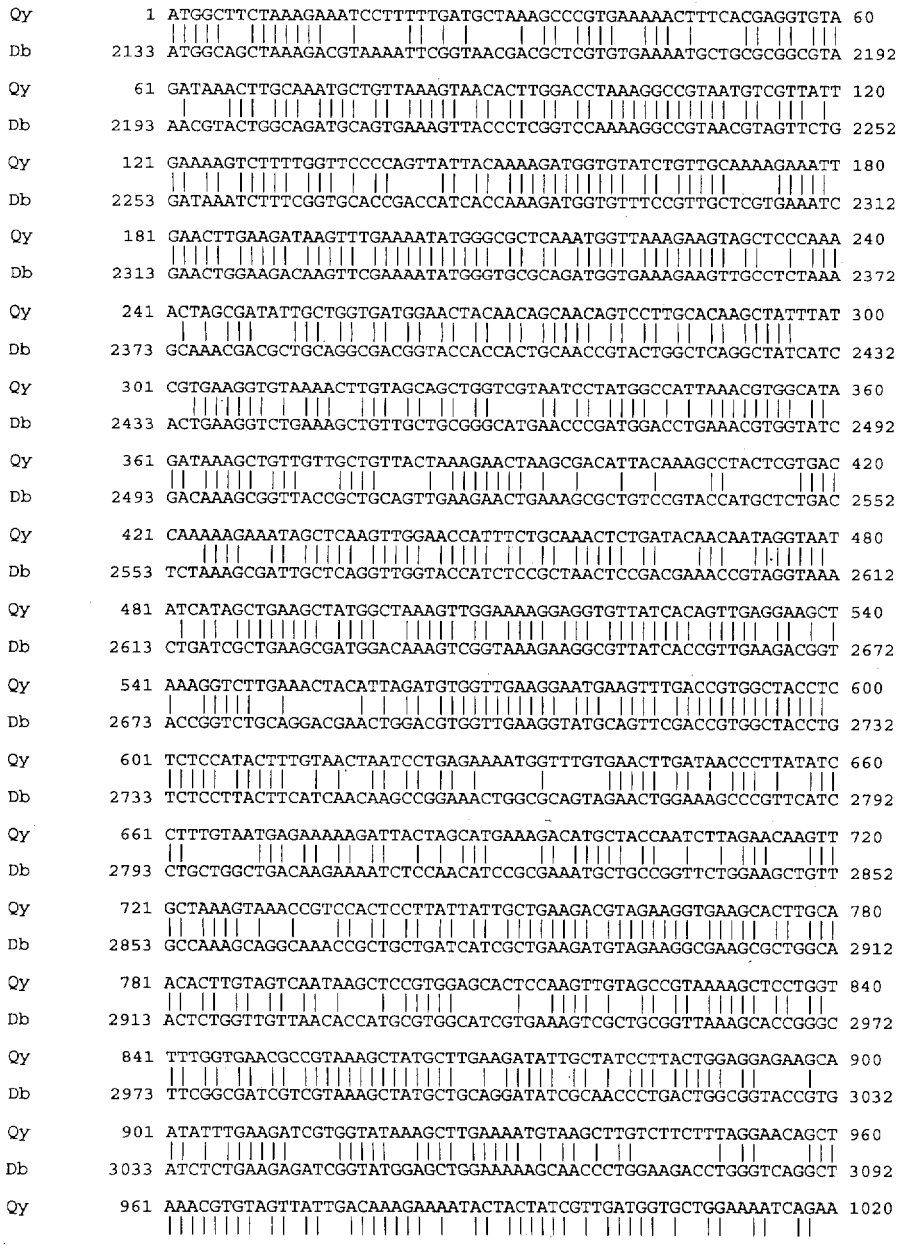


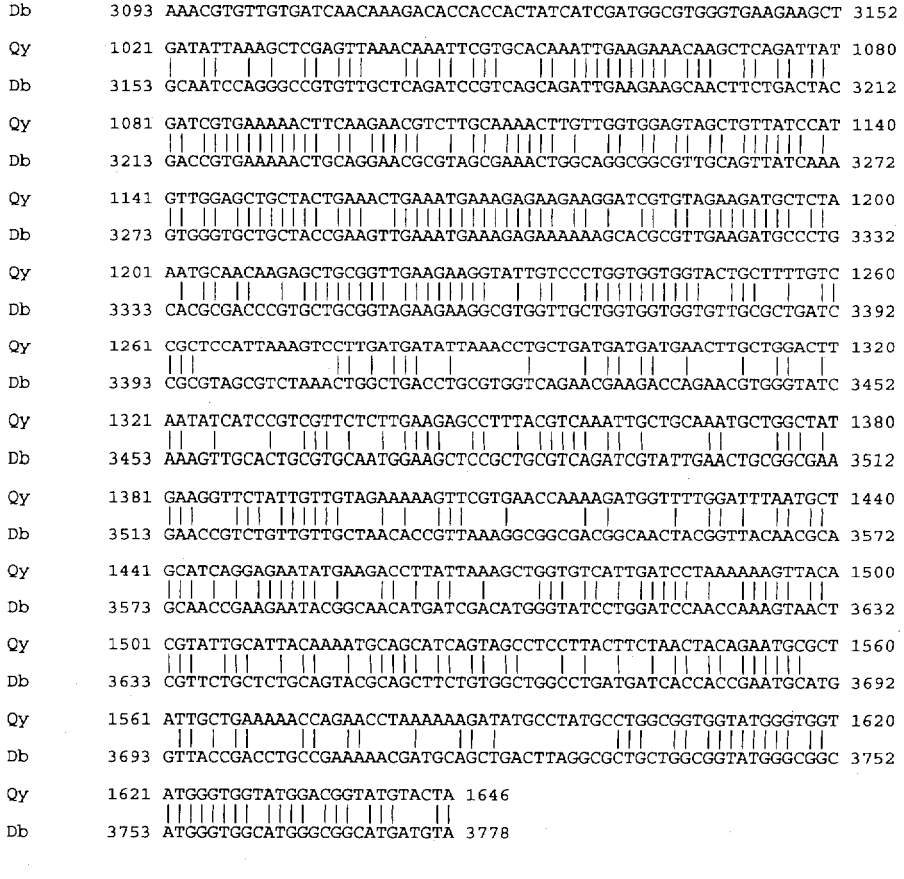


RESULT 15
US-10-157-096-26
; Sequence 26, Application US/10157096
; Publication No. US20030190601A1
; GENERAL INFORMATION:
; APPLICANT: Sabbadini, Roger A.
; APPLICANT: Neil Berkley
; APPLICANT: Surber, Mark W.
; TITLE OF INVENTION: TARGET DISPLAY ON MINICELLS
; FILE REFERENCE: MPEX.008DV12
; CURRENT APPLICATION NUMBER: US/10/157,096
; CURRENT FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/293,566
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/359,843
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 10/154,51
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3840
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Fusion protein
US-10-157-096-26

Query Match 41.6%; Score 684.4; DB 13; Length 3840;
Best Local Similarity 63.5%; Pred. No. 6.6e-130;
Matches 1045; Conservative 0; Mismatches 601; Indels 0; Gaps 0;





Search completed: January 29, 2004, 16:43:39
Job time : 485 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 13:07:12 ; Search time 89 Seconds
(without alignments)
8168.068 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 1647
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

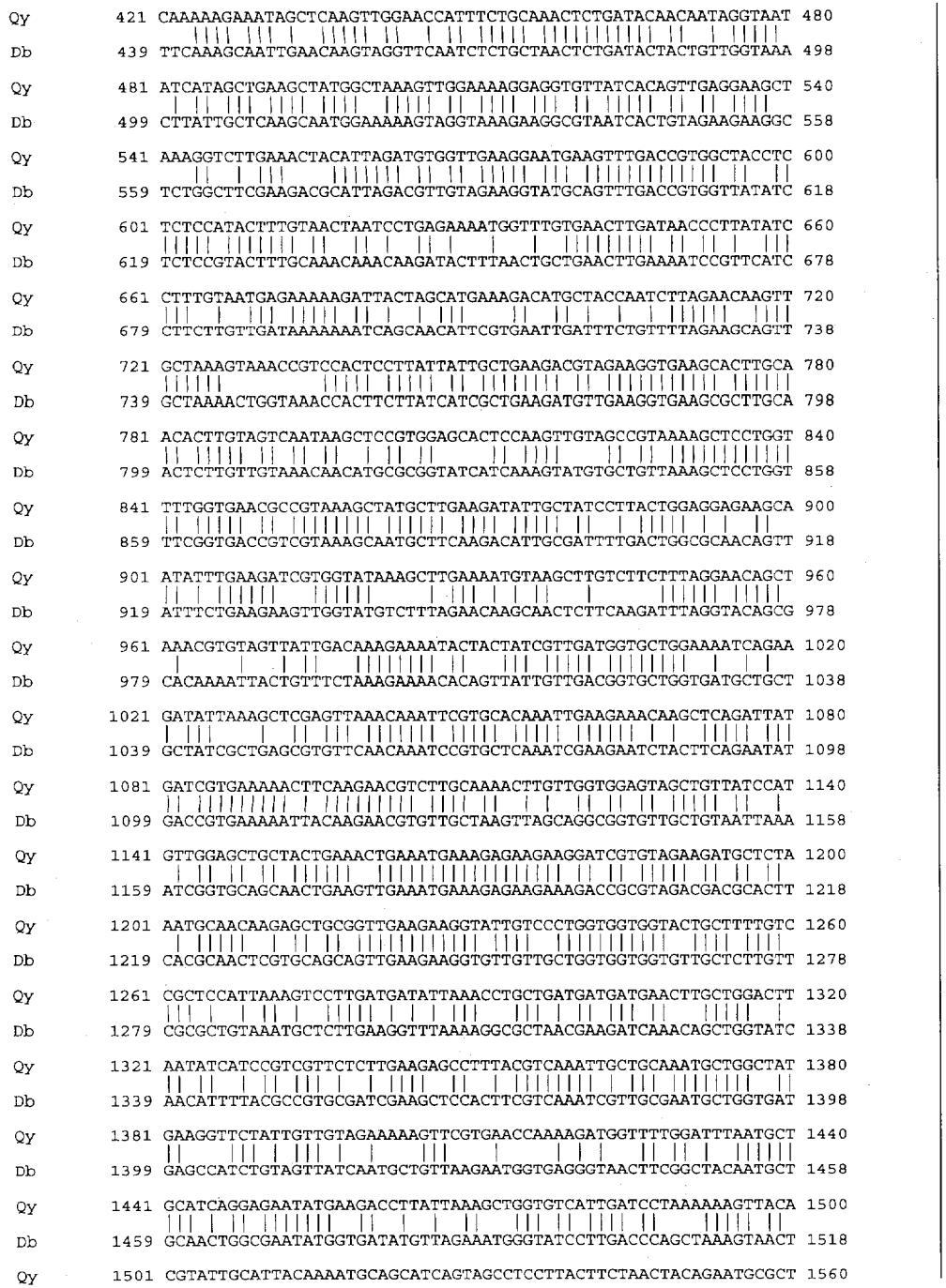
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	748.8	45.5	1653	4	US-09-328-352-1382	Sequence 1382, Ap
	2	725.8	44.1	1830121	4	US-09-557-884-1	Sequence 1, Appli
	3	725.8	44.1	1830121	4	US-09-643-990A-1	Sequence 1, Appli
c	4	690.2	41.9	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	5	689.8	41.9	1926	4	US-09-613-303-50	Sequence 50, Appl
	6	687.6	41.7	5365	4	US-08-961-527-77	Sequence 77, Appl
	7	682.8	41.5	1647	3	US-09-472-971-5	Sequence 5, Appli
	8	682.8	41.5	4524	3	US-09-472-971-7	Sequence 7, Appli
	9	642.4	39.0	1838	3	US-08-470-260-7	Sequence 7, Appli
	10	642.4	39.0	1838	3	US-08-471-491-7	Sequence 7, Appli
	11	642.4	39.0	1838	3	US-08-466-662-7	Sequence 7, Appli
	12	642.4	39.0	1838	4	US-08-256-847C-6	Sequence 6, Appli
	13	640	38.9	1623	4	US-09-134-001C-1868	Sequence 1868, Ap
c	14	626.4	38.0	2416	4	US-09-221-017B-895	Sequence 895, App
	15	610.4	37.1	2284	2	US-08-467-822-28	Sequence 28, Appl
	16	610.4	37.1	2284	3	US-08-432-697-28	Sequence 28, Appl
	17	610.4	37.1	2284	3	US-08-466-248-28	Sequence 28, Appl
	18	553.2	33.6	1656	4	US-09-252-991A-7638	Sequence 7638, Ap
c	19	553.2	33.6	1722	4	US-09-252-991A-7917	Sequence 7917, Ap
	20	480.6	29.2	1050	4	US-09-107-532A-3030	Sequence 3030, Ap
	21	441.2	26.8	1777	4	US-08-828-199A-1	Sequence 1, Appli
c	22	440	26.7	580073	4	US-08-545-528D-1	Sequence 1, Appli
	23	439.8	26.7	2051	4	US-09-207-388-8	Sequence 8, Appli
	24	439.4	26.7	1647	4	US-09-207-388-9	Sequence 9, Appli
	25	439.4	26.7	1707	4	US-09-207-388-10	Sequence 10, Appl
	26	427.4	26.0	1653	4	US-09-207-388-6	Sequence 6, Appli
	27	427.4	26.0	1713	4	US-09-207-388-7	Sequence 7, Appli
	28	425.8	25.9	1944	4	US-09-613-303-52	Sequence 52, Appl
	29	406.2	24.7	1623	4	US-09-613-303-3	Sequence 3, Appli
	30	406.2	24.7	1920	4	US-09-613-303-16	Sequence 16, Appl
	31	406.2	24.7	1947	4	US-09-613-303-28	Sequence 28, Appl
	32	406.2	24.7	2847	4	US-09-613-303-20	Sequence 20, Appl
	33	406.2	24.7	4380	4	US-08-955-565A-3	Sequence 3, Appli
	34	406.2	24.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	35	406.2	24.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	36	396.6	24.1	1620	2	US-08-461-775-10	Sequence 10, Appl
	37	396.6	24.1	1620	3	US-09-031-606-10	Sequence 10, Appl
	38	390.8	23.7	1569	2	US-08-997-080-113	Sequence 113, App
	39	390.8	23.7	1569	2	US-08-997-362-113	Sequence 113, App
	40	390.8	23.7	1569	3	US-09-095-855-113	Sequence 113, App
	41	390.8	23.7	1569	4	US-09-324-542-113	Sequence 113, App
	42	390.8	23.7	1569	4	US-09-205-426-113	Sequence 113, App
	43	390.8	23.7	1626	2	US-08-997-080-159	Sequence 159, App
	44	390.8	23.7	1626	2	US-08-997-362-159	Sequence 159, App
	45	390.8	23.7	1626	3	US-09-095-855-159	Sequence 159, App

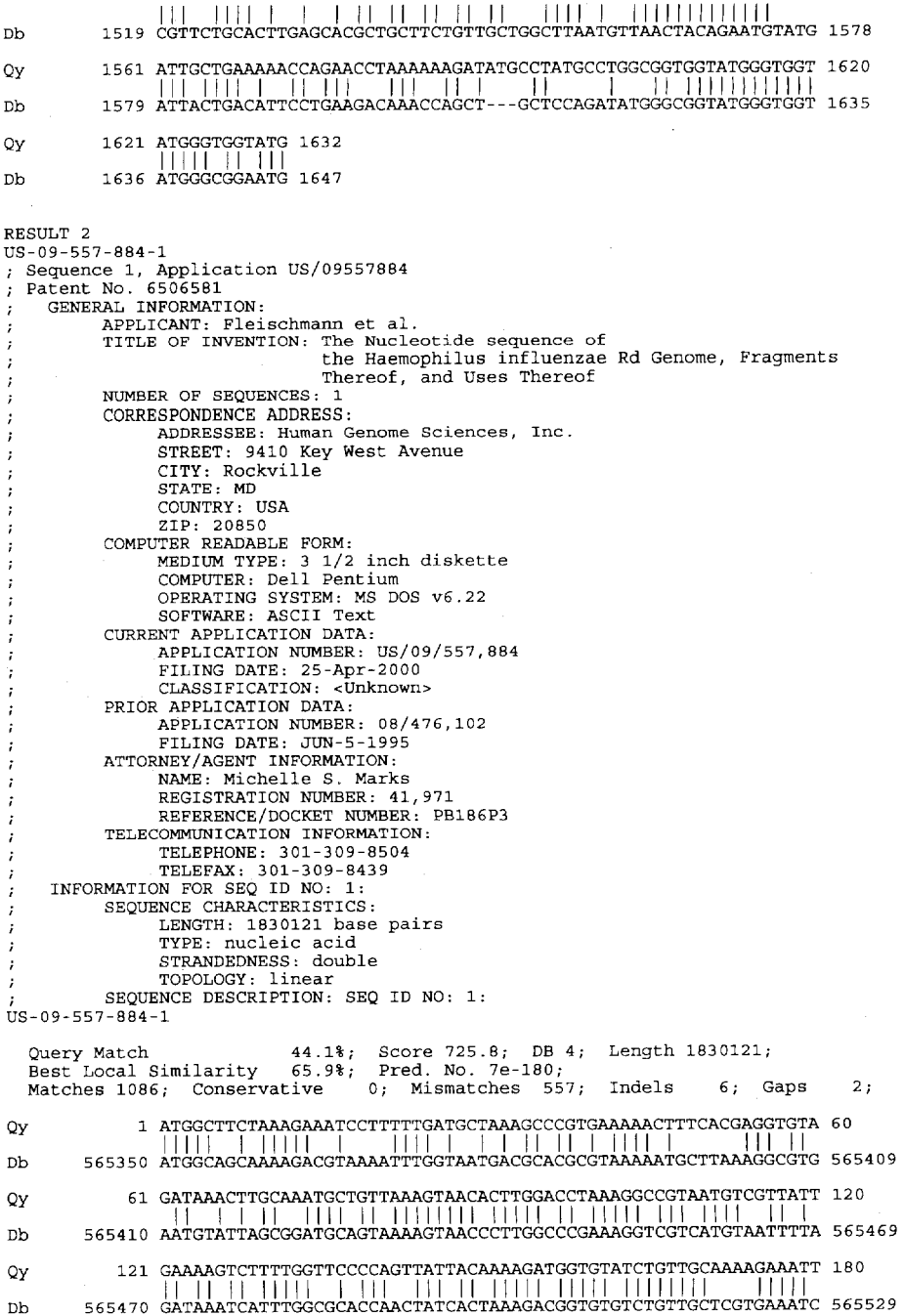
ALIGNMENTS

RESULT 1
US-09-328-352-1382
; Sequence 1382, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1382
; LENGTH: 1653
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1382

Query Match 45.5%; Score 748.8; DB 4; Length 1653;
Best Local Similarity 66.6%; Pred. No. 2.2e-187;
Matches 1087; Conservative 0; Mismatches 542; Indels 3; Gaps 1;

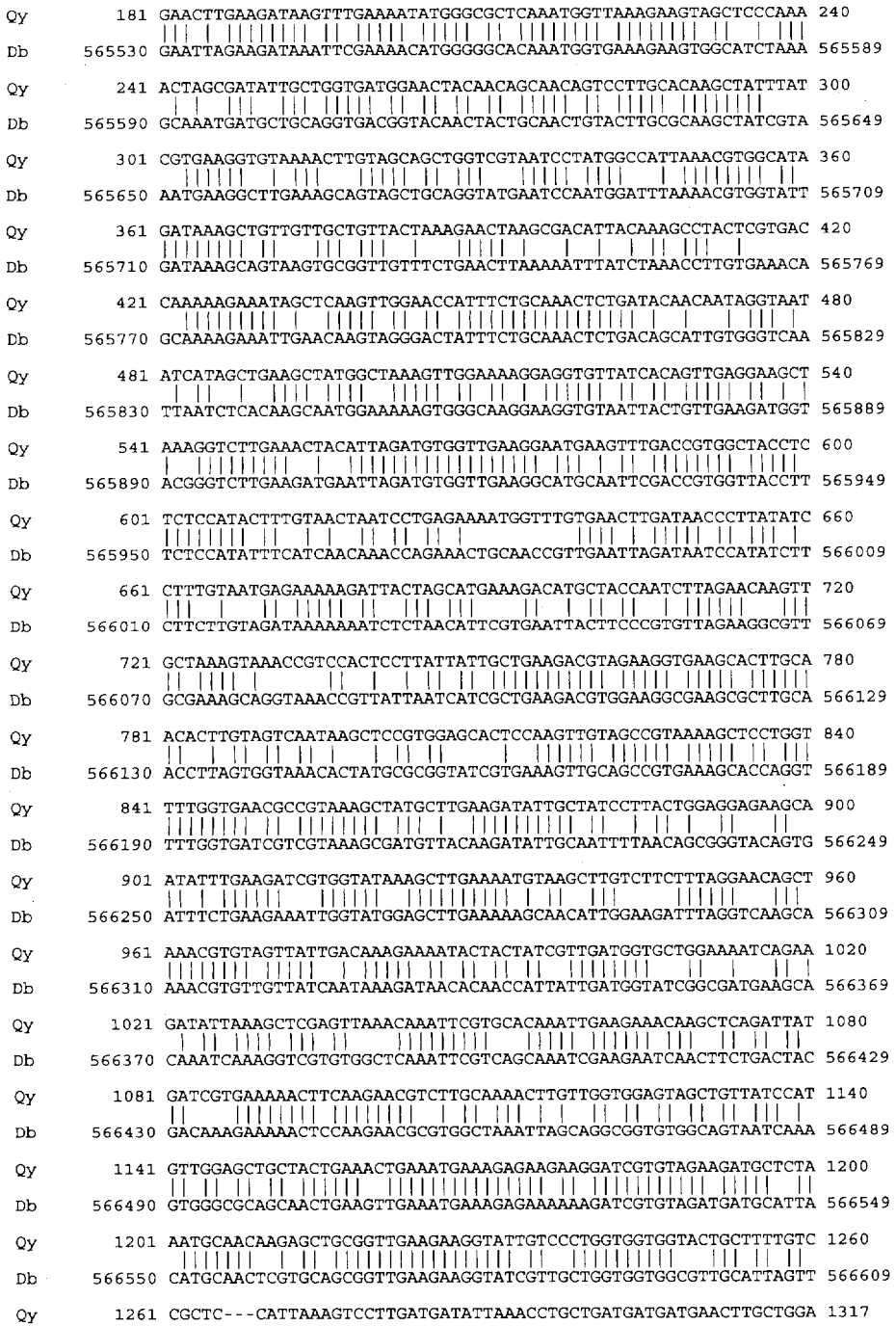
```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||| |  |||||||  |    |||| || |  ||| ||   |||| |      ||| |||
Db         19 ATGTCAGCTAAAGACGTAAAATTTGGTGATTCAGCTCGCTCAAAAATGATTGCAGGGGTA 78

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                |  ||||||| |||| |||||||| ||  | || ||||||||||||||||| || ||
Db         79 AACGTACTTGCAGATGCGGTTAAAGTGACTTTAGGTCCTAAAGGCCGTAATGTTGTGATC 138

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
               ||    |||| ||  | ||    || || ||||| |||||| ||||||||||||||||
Db        139 GACCGCTCTTTCGGCGCGCCGCACATCACTAAAGACGGTGTAACTGTTGCAAAAGAAATT 198

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
               | |  |||| |||||||||| ||||| |||||| | |||   |||||  |   ||||
Db        199 TCATTAAAAGACAAGTTTGAAAACATGGGTGCTCAACTTGTTCGTGAAGTTTCAAGCAAA 258

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||| ||| || || ||||| || || ||||| ||||| || ||||| ||||| ||||
Db        259 ACTAACGACATCGCAGGTGACGGTACAACAACTGCAACTGTACTTGCTCAAGCAATTTTA 318

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||| | |||   ||| | ||||||   || || ||||   | ||||| || ||
Db        319 AATGAAGGTATCAAATCAGTAACTGCTGGTATGAACCCAATGGATTTAAAACGCGGTATC 378

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              || |  || ||    ||||  | || |  |  |   ||| | || ||  ||  |||
Db        379 GACATTGCAGTAAAAACTGTAGTTGAAAATATCCGTTCTATTGCTAAACCAGCTGATGAT 438
```

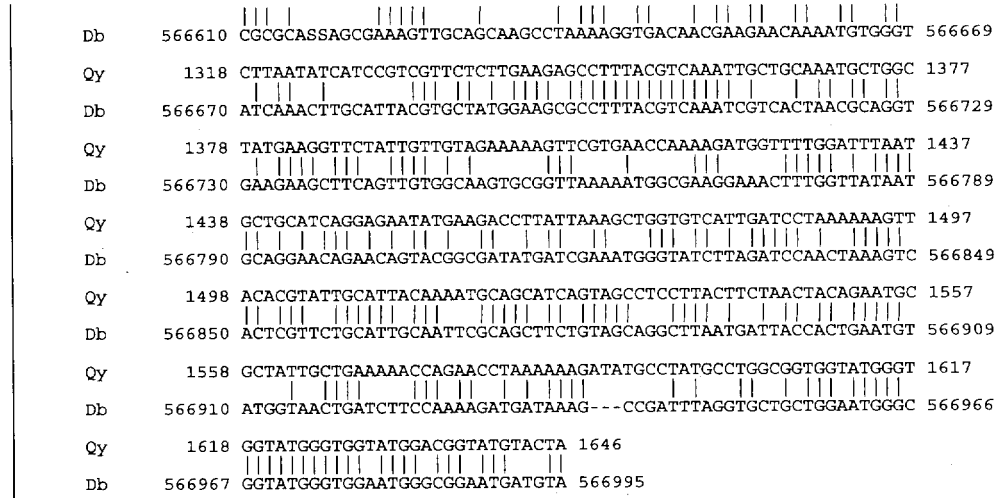




RESULT 2
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 44.1%; Score 725.8; DB 4; Length 1830121;
Best Local Similarity 65.9%; Pred. No. 7e-180;
Matches 1086; Conservative 0; Mismatches 557; Indels 6; Gaps 2;





RESULT 3
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs

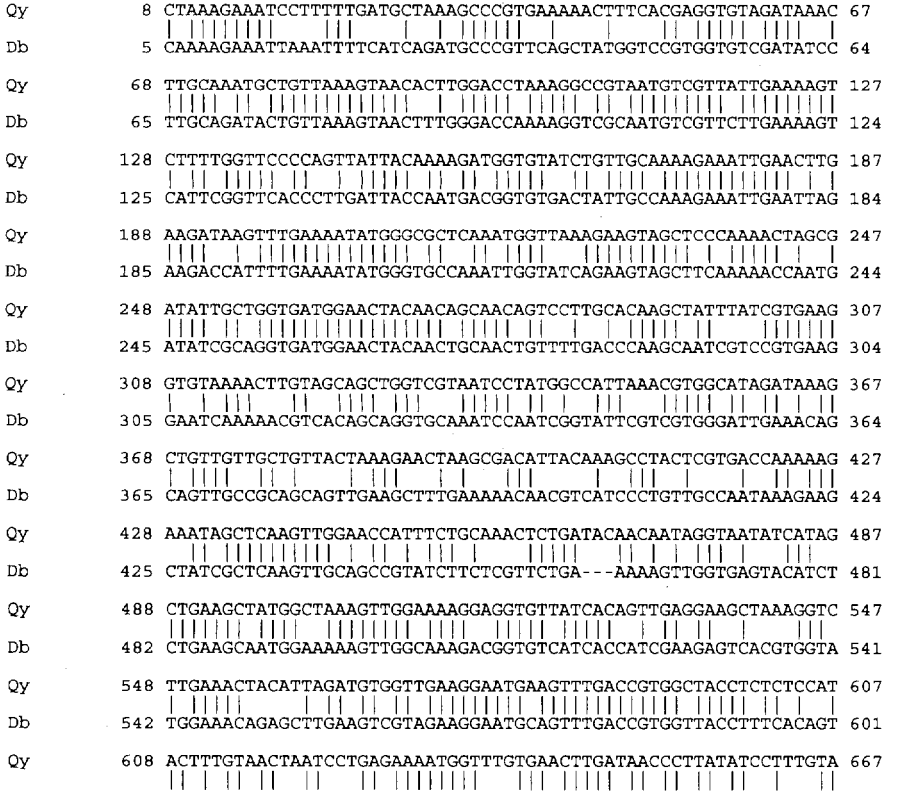
```
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match            44.1%;  Score 725.8;  DB 4;  Length 1830121;
  Best Local Similarity  65.9%;  Pred. No. 7e-180;
  Matches 1086;  Conservative    0;  Mismatches  557;  Indels    6;  Gaps    2;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db     565350 ATGGCAGCAAAAGACGTAAAATTTGGTAATGACGCACGCGTAAAAATGCTTAAAGGCGTG 565409

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db     565410 AATGTATTAGCGGATGCAGTAAAAGTAACCCTTGGCCCGAAAGGTCGTCATGTAATTTTA 565469

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db     565470 GATAAATCATTTGGCGCACCAACTATCACTAAAGACGGTGTGTCTGTTGCTCGTGAAATC 565529

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db     565530 GAATTAGAAGATAAATTCGAAAACATGGGGGCACAAATGGTGAAAGAAGTGGCATCTAAA 565589

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db     565590 GCAAATGATGCTGCAGGTGACGGTACAACTACTGCAACTGTACTTGCGCAAGCTATCGTA 565649

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db     565650 AATGAAGGCTTGAAAGCAGTAGCTGCAGGTATGAATCCAATGGATTTAAAACGTGGTATT 565709

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db     565710 GATAAAGCAGTAAGTGCGGTTGTTTCTGAACTTAAAAATTTATCTAAACCTTGTGAAACA 565769

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db     565770 GCAAAAGAAATTGAACAAGTAGGGACTATTTCTGCAAACTCTGACAGCATTGTGGGTCAA 565829

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db     565830 TTAATCTCACAAGCAATGGAAAAAGTGGGCAAGGAAGGTGTAATTACTGTTGAAGATGGT 565889

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db     565890 ACGGGTCTTGAAGATGAATTAGATGTGGTTGAAGGCATGCAATTCGACCGTGGTTACCTT 565949

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db     565950 TCTCCATATTTCATCAACAAACCAGAAACTGCAACCGTTGAATTAGATAATCCATATCTT 566009

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db     566010 CTTCTTGTAGATAAAAAAATCTCTAACATTCGTGAATTACTTCCCGTGTTAGAAGGCGTT 566069

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db     566070 GCGAAAGCAGGTAAACCGTTATTAATCATCGCTGAAGACGTGGAAGGCGAAGCGCTTGCA 566129

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db     566130 ACCTTAGTGGTAAACACTATGCGCGGTATCGTGAAAGTTGCAGCCGTGAAAGCACCAGGT 566189

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db     566190 TTTGGTGATCGTCGTAAAGCGATGTTACAAGATATTGCAATTTTAACAGCGGGTACAGTG 566249

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db     566250 ATTTCTGAAGAAATTGGTATGGAGCTTGAAAAAGCAACATTGGAAGATTTAGGTCAAGCA 566309

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db     566310 AAACGTGTTGTTATCAATAAAGATAACACAACCATTATTGATGGTATCGGCGATGAAGCA 566369

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db     566370 CAAATCAAAGGTCGTGTGGCTCAAATTCGTCAGCAAATCGAAGAATCAACTTCTGACTAC 566429

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db     566430 GACAAAGAAAAACTCCAAGAACGCGTGGCTAAATTAGCAGGCGGTGTGGCAGTAATCAAA 566489

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db     566490 GTGGGCGCAGCAACTGAAGTTGAAATGAAAGAGAAAAAAGATCGTGTAGATGATGCATTA 566549

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db     566550 CATGCAACTCGTGCAGCGGTTGAAGAAGGTATCGTTGCTGGTGGTGGCGTTGCATTAGTT 566609

Qy       1261 CGCTC---CATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db     566610 CGCGCASSAGCGAAAGTTGCAGCAAGCCTAAAAGGTGACAACGAAGAACAAAATGTGGGT 566669

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db     566670 ATCAAACTTGCATTACGTGCTATGGAAGCGCCTTTACGTCAAATCGTCACTAACGCAGGT 566729

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437
Db     566730 GAAGAAGCTTCAGTTGTGGCAAGTGCGGTTAAAAATGGCGAAGGAAACTTTGGTTATAAT 566789

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db     566790 GCAGGAACAGAACAGTACGGCGATATGATCGAAATGGGTATCTTAGATCCAACTAAAGTC 566849

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db     566850 ACTCGTTCTGCATTGCAATTCGCAGCTTCTGTAGCAGGCTTAATGATTACCACTGAATGT 566909

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db     566910 ATGGTAACTGATCTTCCAAAAGATGATAAAG---CCGATTTAGGTGCTGCTGGAATGGGC 566966

Qy       1618 GGTATGGGTGGTATGGACGGTATGTACTA 1646
Db     566967 GGTATGGGTGGAATGGGCGGAATGATGTA 566995

RESULT 4
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
;   LENGTH: 1230025
;   TYPE: DNA
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(15000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (15001)..(30000)
```

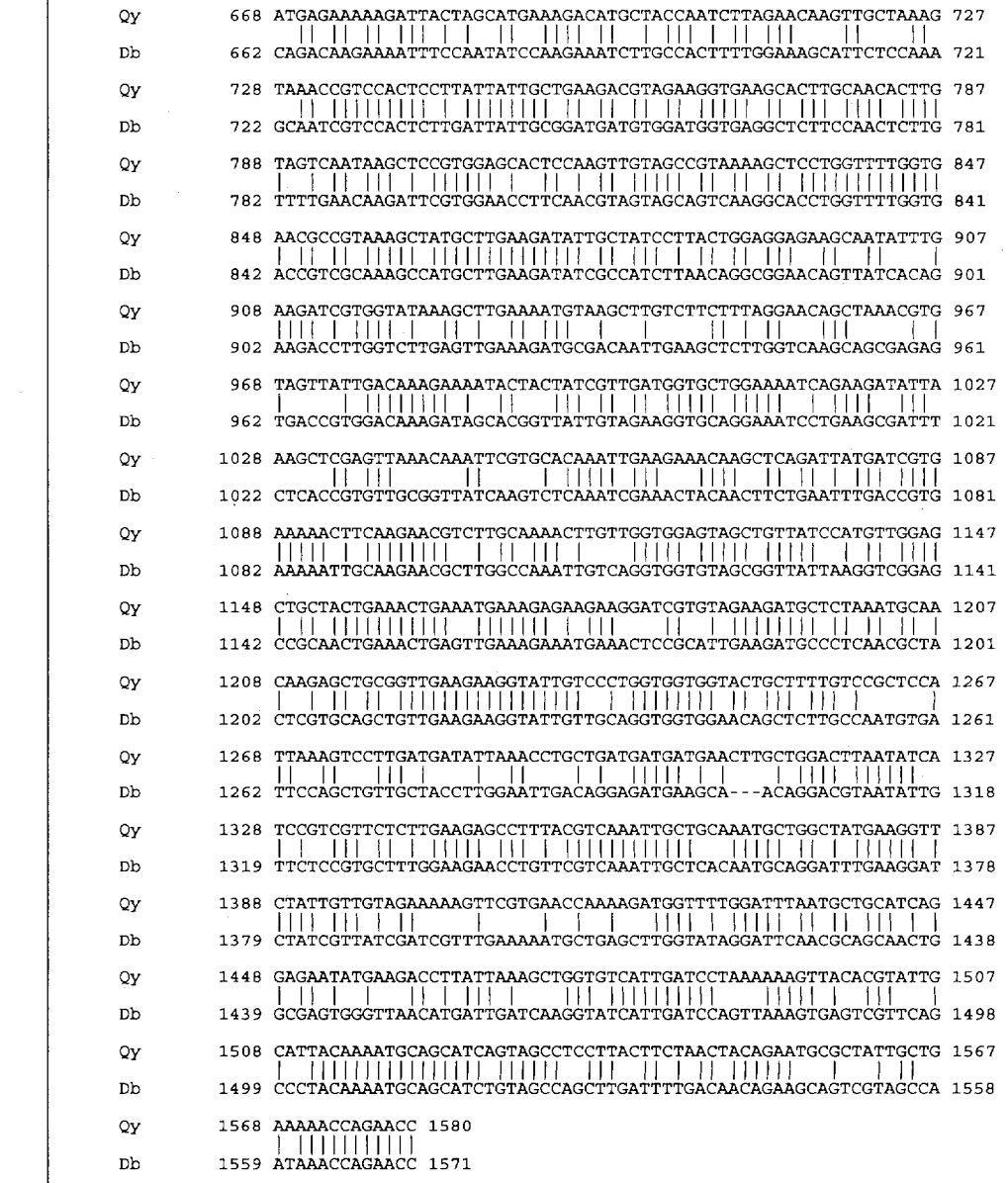
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

```
;    LOCATION: (750001)..(765000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (765001)..(780000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (780001)..(795000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (795001)..(810000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (810001)..(825000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (825001)..(840000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (840001)..(855000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (855001)..(870000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (870001)..(885000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (885001)..(900000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (900001)..(915000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature

  Query Match             41.9%;  Score 690.2;  DB 4;  Length 1230025;
  Best Local Similarity   65.1%;  Pred. No. 1.3e-170;
  Matches 1035;  Conservative    0;  Mismatches  549;  Indels    6;  Gaps    1;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db     179472 ATGGCAGCGAAAAATATTAAATATAATGAAGAAGCCAGANAAAAAATACATAAAGGGGTA 179413

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db     179412 AAAACTCTTGCAGAAGCAGTAAAAGTTACTCTAGGTCCTAAAGGACGTCACGTAGTTATA 179353

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db     179352 GATAAGAGCTTTGGCTCTCCCCAAGTGACTAAAGATGGTGTTACTGTAGCTAAAGAAATC 179293

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db     179292 GAGCTCGAAGACAAACATGAAAACATGGGCGCTCAGATGGTAAAAGAAGTCGCCAGCAAA 179233

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db     179232 ACTGCTGACAAAGCAGGCGACGGAACTACAACAGCAACTGTTCTTGCAGAAGCAATCTAT 179173

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db     179172 AGCGAAGGTCTAAGAAATGTCACTGCCGGTGCCAATCCTATGGACCTAAAAAGAGGTATC 179113

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db     179112 GACAAAGCCGTAAAAGTTGTTGTTGATGAACTCAAAAAAATTAGTAAACCTGTACAACAT 179053

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db     179052 CACAAAGAAATCGCTCAAGTAGCTACTATCTCAGCAAATAATGATTCCGAAATCGGAAAT 178993

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db     178992 CTTATTGCAGAAGCTATGGAAAAAGTTGGTAAAAACGGATCCATTACTGTTGAAGAAGCT 178933

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db     178932 AAAGGCTTCGAAACTGTTCTCGACGTTGTAGAAGGAATGAACTTCAACCGTGGATACCTC 178873

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db     178872 TCCAGCTACTTCTCCACAAATCCAGAAACTCAAGAATGCGTTTTAGAAGACGCTCTGATT 178813

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db     178812 CTAATCTACGATAAAAAAATCTCTGGAATTAAAGACTTCCTTCCAGTTTTACAACAAGTA 178753

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db     178752 GCAGAATCTGGACGCCCTCTTTTAATCATTGCAGAAGAAATTGAAGGAGAAGCTTTAGCA 178693

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db     178692 ACTCTAGTAGTCAATAGACTCCGTGCAGGATTCAGAGTCTGTGCAGTGAAAGCTCCTGGT 178633

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db     178632 TTCGGTGACAGAAGAAAAGCTATGTTAGAAGACATCGCTATCCTTACTGGTGGCCAACTA 178573

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db     178572 GTTAGCGAAGAACTTGGCATGAAACTAGAGAATACAACTCTAGCAATGTTAGGAAAAGCT 178513

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db     178512 AAGAAAGTTATCGTAACTAAAGAAGATACCACAATCGTCGAAGGCTTAGGAAACAAACCT 178453

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db     178452 GATATCCAAGCTCGATGCGACAATATTAAAAAACAAATCGAAGATAGCACTTCAGATTAC 178393

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db     178392 GACAAAGAAAAACTCCAAGAGCGTTTAGCTAAACTCTCCGGTGGTGTCGCCGTAATCCGC 178333

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db     178332 GTAGGAGCTGCTACCGAAATAGAGATGAAAGAGAAAAAAGACAGAGTAGATGATGCACAA 178273

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db     178272 CACGCAACCATTGCAGCTGTCGAAGAAGGAATCCTCCCTGGTGGTGGAACTGCCTTAGTT 178213

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCT------GCTGATGATGATGAACTTGCT 1314
Db     178212 CGCTGTATCCCTACACTAGAAGCTTTCCTTCCTATGCTAGCAAACGAAGACGAAGCTATT 178153

Qy       1315 GGACTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCT 1374
Db     178152 GGTACTCGTATTATTCTAAAAGCATTAACAGCTCCATTAAAGCAAATTGCAAGTAACGCA 178093

Qy       1375 GGCTATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTT 1434
Db     178092 GGTAAAGAAGGCGCTATCATTTGTCAGCAAGTTCTAGCAAGATCTGCAAATGAAGGCTAT 178033

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db     178032 GATGCTTTACGTGACGCTTATACAGATATGATTGACGCAGGAATTTTAGATCCAACTAAA 177973

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db     177972 GTGACTCGCTCAGCTCTAGAAAGCGCAGCTTCTATCGCAGGATTACTCCTCACAACAGAA 177913

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAA 1584
Db     177912 GCCTTAATCGCTGATATCCCAGAAGAGAAA 177883
```

RESULT 5
US-09-613-303-50
; Sequence 50, Application US/09613303
; Patent No. 6495347
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/09/613,303
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
; NAME/KEY: CDS
; LOCATION: (1)...(1923)
US-09-613-303-50

Query Match 41.9%; Score 689.8; DB 4; Length 1926;
Best Local Similarity 65.8%; Pred. No. 7.7e-172;
Matches 1035; Conservative 0; Mismatches 532; Indels 6; Gaps 2;





RESULT 6
US-08-961-527-77
; Sequence 77, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch

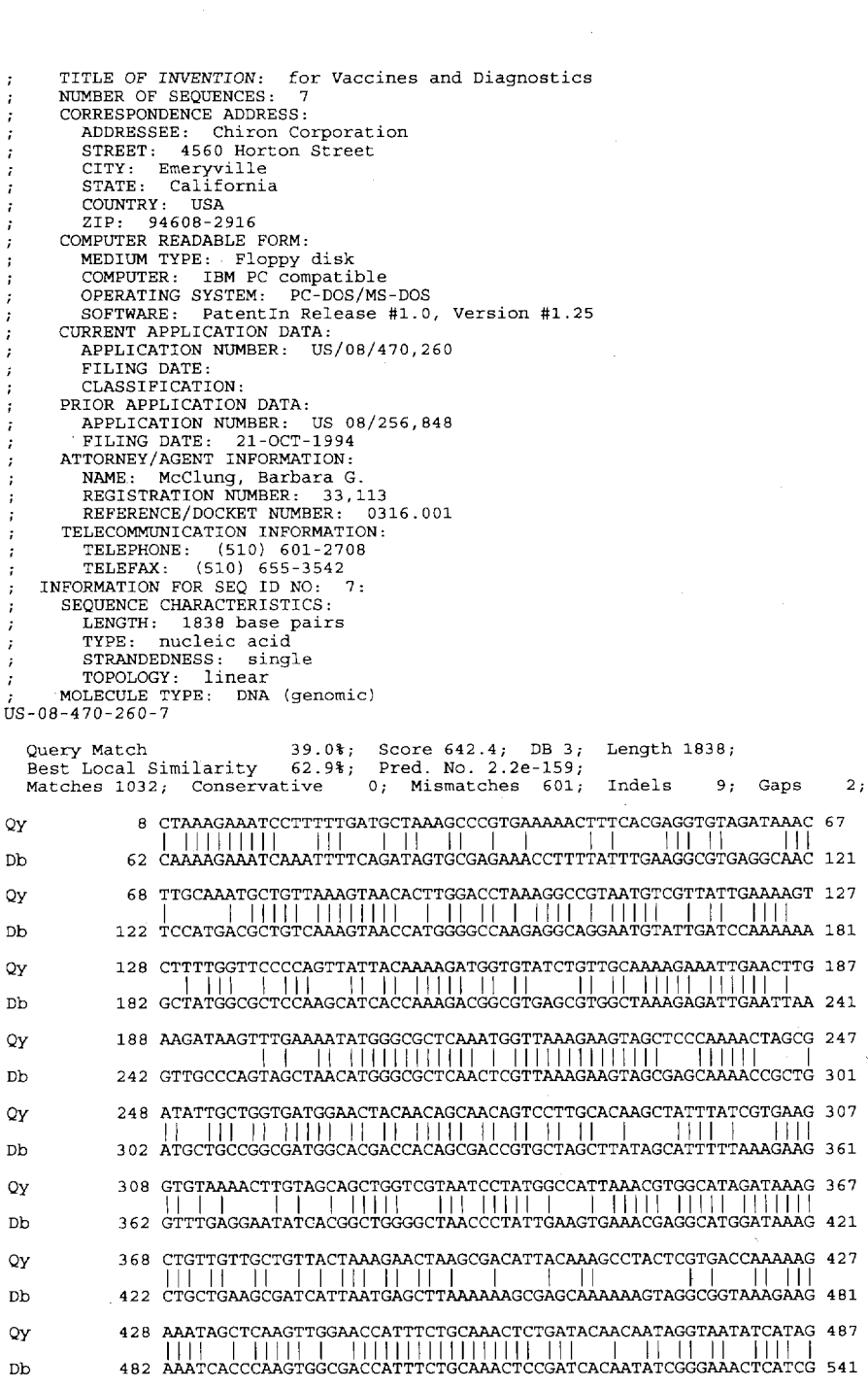
```
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  77:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5365 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-77

  Query Match             41.7%;  Score 687.6;  DB 4;  Length 5365;
  Best Local Similarity   65.7%;  Pred. No. 4.8e-171;
  Matches 1034;  Conservative    0;  Mismatches  534;  Indels    6;  Gaps    2;

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db        281 TCAAAAGAAATTAAATTTTCATCAGATGCCCGTTCAGCCATGGTTCGTGGTGTCGATATC 340

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db        341 CTTGCAGACACTGTTAAAGTAACCTTGGGACCAAAAGGTCGCAATGTCGTTCTTGAAAAG 400

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db        401 TCATTCGGTTCACCCTTGATTACCAATGACGGTGTGACCATTGCCAAAGAAATCGAATTG 460

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db        461 GAAGACCATTTTGAAAATATGGGTGCTAAGTTAGTATCAGAAGTAGCTTCTAAAACCAAT 520

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTATCGTGAA 306
Db        521 GATATCGCAGGTGACGGAACTACGACTGCAACAGTCTTGACCCAAGCTATCGTCCGTGAA 580

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        581 GGAATCAAAAACGTCACAGCAGGTGCAAATCCAATCGGTATTCGTCGTGGGATTGAAACA 640

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        641 GCAGTTGCCGCAGCAGTTGAAGCTTTGAAAAACAACGCCATCCCTGTTGCCAATAAAGAA 700

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        701 GCTATCGCTCAAGTTGCAGCCGTATCTTCTCGTTCTGA---AAAAGTTGGTGAGTACATC 757

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        758 TCTGAAGCAATGGAAAAAGTTGGCAAAGACGGTGTCATCACCATCGAAGAGTCACGTGGT 817

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db        818 ATGGAAACAGAGCTTGAAGTCGTAGAAGGAATGCAGTTTGACCGTGGTTACCTTTCACAG 877

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        878 TACATGGTGACAGATAGCGAAAAAATGGTGGCTGACCTTGAAAATCCGTACATTTTGATT 937

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        938 ACAGACAAGAAAATTTCCAATATCCAAGAAATCTTGCCACTTTTGGAAAGCATTCTCCAA 997

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        998 AGCAATCGTCCACTCTTGATTATTGCGGATGATGTGGATGGCGAGGCTCTTCCAACTCTT 1057

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db       1058 GTTTTGAACAAGATTCGTGGAACCTTCAACGTAGTAGCAGTCAAGGCACCTGGTTTTGGT 1117

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db       1118 GACCGTCGCAAAGCCATGCTTGAAGATATCGCCATCTTAACAGGCGGAACAGTTATCACA 1177

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db       1178 GAAGACCTTGGTCTTGAGTTGAAAGATGCGACAATTGAAGCTCTTGGTCAAGCAGCGAGA 1237

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db       1238 GTGACCGTGGACAAAGATAGCACGGTTATTGTAGAAGGTGCAGGAAATCCTGAAGCGATT 1297

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db       1298 TCTCACCGTGTTGCGGTTATCAAGTCTCAAATCGAAACTACAACTTCTGAATTTGACCGT 1357

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db       1358 GAAAAATTGCAAGAACGCTTGGCCAAATTGTCAGGTGGTGTAGCGGTTATTAAGGTTGGA 1417

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db       1418 GCCGCAACTGAAACTGAGTTGAAAGAAATGAAACTCCGCATTGAAGATGCCCTCAACGCT 1477

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db       1478 ACTCGTGCAGCTGTTGAAGAAGGTATTGTTGCAGGTGGTGGAACAGCTCTTGCCAATGTG 1537

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db       1538 ATTCCAGCTGTTGCTACCTTGGAATTGACAGGAGATGAAGCA---ACAGGACGTAATATT 1594

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db       1595 GTTCTCCGTGCTTTGGAAGAACCCGTTCGTCAAATTGCTCACAATGCAGGATTTGAAGGA 1654

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db       1655 TCTATCGTTATCGATCGTTTGAAAAATGCTGAGCTTGGTATAGGATTTAACGCAGCAACT 1714

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db       1715 GGCGAGTGGGTTAACATGATTGATCAAGGTATCATTGATCCAGTTAAAGTGAGTCGTTCA 1774

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db       1775 GCCCTACAAAATGCAGCATCTGTAGCCAGCTTGATTTTGACAACAGAAGCAGTCGTAGCC 1834

Qy       1567 GAAAAACCAGAACC 1580
```

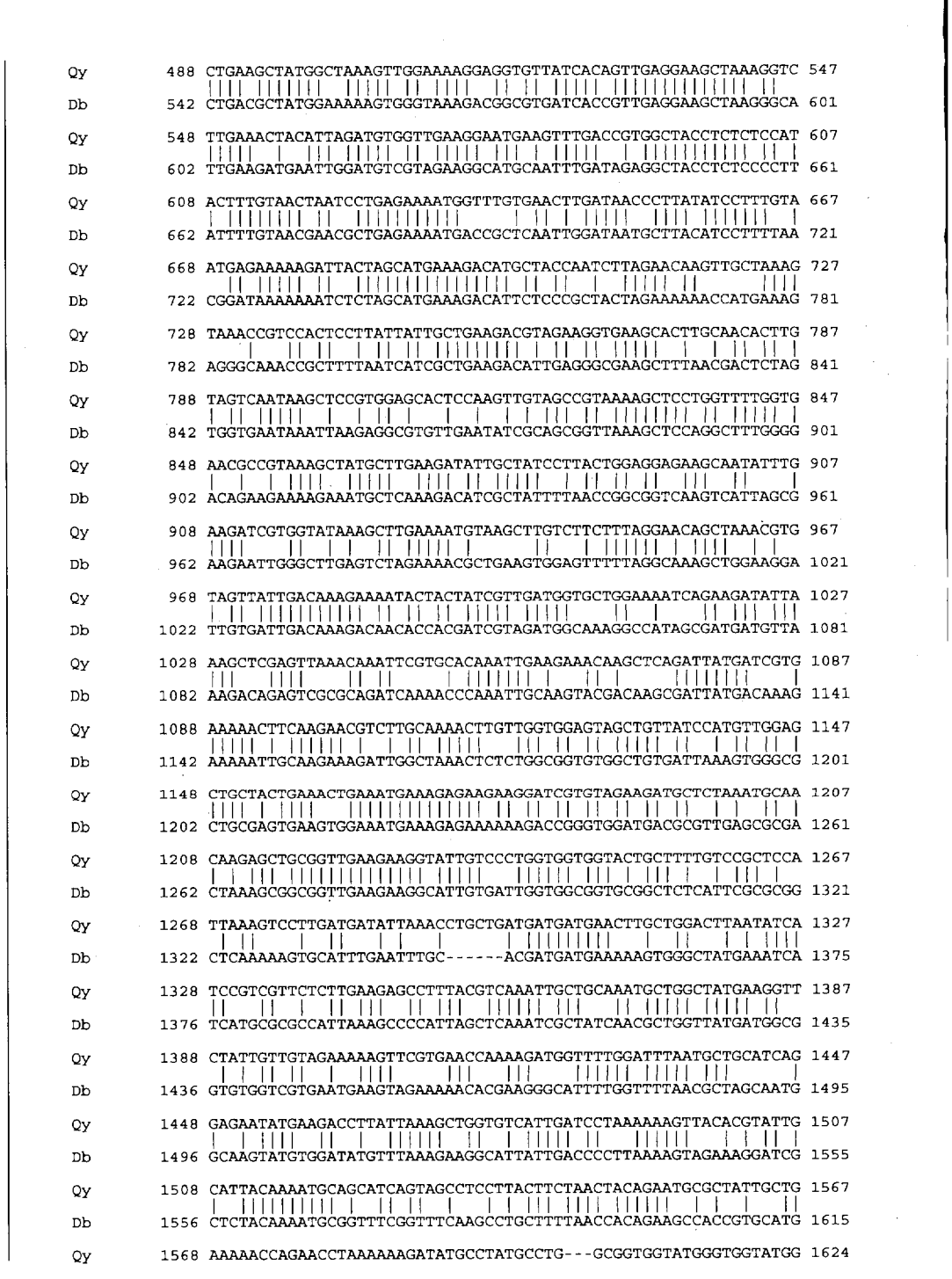
```
Db       1835 AATAAACCAGAACC 1848


RESULT 7
US-09-472-971-5
; Sequence 5, Application US/09472971
; Patent No. 6197547
; GENERAL INFORMATION:
;  APPLICANT: SOGO, Kazuyo
;  APPLICANT: YANAGI, Hideki
;  APPLICANT: YURA, Takashi
;  TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
;  FILE REFERENCE: 1422-409P
;  CURRENT APPLICATION NUMBER: US/09/472,971
;  CURRENT FILING DATE: 1999-12-28
;  EARLIER APPLICATION NUMBER: JP10-372965
;  EARLIER FILING DATE: 1998-12-28
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1647
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-472-971-5

  Query Match             41.5%;  Score 682.8;  DB 3;  Length 1647;
  Best Local Similarity   63.4%;  Pred. No. 5e-170;
  Matches 1044;  Conservative    0;  Mismatches  602;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              |||||  |||||||  |    || | |    | || ||||  ||| |    || || |||
Db          1 ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA 60

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                |   ||| ||| |||| || |||||  || || || || ||||||||||| || |||  |
Db         61 AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG 120

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              || || ||||| ||| | ||    || || ||||||||||| || |||||    |||||
Db        121 GATAAATCTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC 180

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||| ||||| ||||| |||||||||||| || || ||||| |||||||| ||  | |||
Db        181 GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA 240

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              | | |||   ||| || || || || || ||||| || || || || |||||
Db        241 GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC 300

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
                |||||| | |||  ||| || || ||   || || |||| | | |||||||| ||
Db        301 ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC 360

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              || |||||  |||   ||||    | ||||||| | |   |   |    ||       ||||
Db        361 GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC 420

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||  || ||||| ||||| |||| || || ||||| ||   ||| ||||||| ||||||
Db        421 TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 480

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              | || |||||||| |||| ||||| || |||| ||| |||||||| ||||| || | |
Db        481 CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT 540

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |  |||||  |      | | || |||||||||||| ||| |||| ||||||||||||||
Db        541 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG 600

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||| |||||  | |  || || || |     |     |||||| || | ||| |  |||
Db        601 TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC 660

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||      ||| || || ||  | | |||     || ||||| ||  |  | |||   |||
Db        661 CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT 720

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              || |||| |  |   || || || || || |||||||| |||||||| ||||| || |||
Db        721 GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA 780

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||  ||| || || |  | |||||     | |||| |  || || ||||| || ||
Db        781 ACTGCTGTTGTTAACACCATTCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC 840

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              || || || |||||||||||||  | ||||| || | ||| |||||  ||     |
Db        841 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG 900

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              || | ||||||  |||||  |||| ||||| | || | ||     | ||   |||
Db        901 ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT 960

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||||||| || ||  ||||||| | || |||||| | |||||  |  ||  ||  ||
Db        961 AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT 1020

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              | || | |  || ||| || || ||| || |||||||| ||| || || ||
Db       1021 GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC 1080

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              || |||||||||| || |||||  | || ||||| |  || || || || |||||| |
Db       1081 GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA 1140

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              || || |||||||| |||  |||||||||||||||| || |  || || |||||||| ||
Db       1141 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG 1200

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              | || ||  | |||||||| ||||||||  | ||  |||||||||||  |||  |  ||
Db       1201 CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC 1260

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||                 || | |||  |     | || ||  |  | || |  |
Db       1261 CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC 1320

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||  |    | ||| | | |||| ||  | ||||| || |     ||     ||| |
Db       1321 AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA 1380

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   ||| ||||||  | |  |||  |     | |     | || | || ||
Db       1381 GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA 1440

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||| | | |||||  |   || | ||  |  ||| || | |||||  ||
Db       1441 GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT 1500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||  |||  | ||  | ||||| || || ||  | |  |  | || || ||||||
Db       1501 CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG 1560

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              || | ||   ||  ||    |  || |          |||  || |||||||| ||
Db       1561 GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC 1620

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
              |||||||| |||| ||| |||    ||
Db       1621 ATGGGTGGCATGGGCGGCATGATGTA 1646


RESULT 8
US-09-472-971-7
; Sequence 7, Application US/09472971
; Patent No. 6197547
```

```
; GENERAL INFORMATION:
;  APPLICANT: SOGO, Kazuyo
;  APPLICANT: YANAGI, Hideki
;  APPLICANT: YURA, Takashi
;  TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
;  FILE REFERENCE: 1422-409P
;  CURRENT APPLICATION NUMBER: US/09/472,971
;  CURRENT FILING DATE: 1999-12-28
;  EARLIER APPLICATION NUMBER: JP10-372965
;  EARLIER FILING DATE: 1998-12-28
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 4524
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-472-971-7

  Query Match            41.5%;  Score 682.8;  DB 3;  Length 4524;
  Best Local Similarity  63.4%;  Pred. No. 8.1e-170;
  Matches 1044;  Conservative    0;  Mismatches  602;  Indels    0;  Gaps    0;

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db        383 ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA 442

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db        443 AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG 502

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db        503 GATAAATCTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC 562

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db        563 GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA 622

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db        623 GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC 682

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        683 ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC 742

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        743 GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC 802

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        803 TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA 862

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        863 CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT 922

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        923 ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG 982

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        983 TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC 1042

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db       1043 CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT 1102

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db       1103 GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA 1162

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db       1163 ACTGCTGTTGTTAACACCATTCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC 1222

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db       1223 TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG 1282

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db       1283 ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT 1342

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db       1343 AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT 1402

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db       1403 GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC 1462

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db       1463 GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA 1522

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db       1523 GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG 1582

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db       1583 CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC 1642

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db       1643 CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC 1702

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db       1703 AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA 1762

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db       1763 GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA 1822

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db       1823 GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT 1882

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db       1883 CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG 1942

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db       1943 GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC 2002

Qy       1621 ATGGGTGGTATGGACGGTATGTACTA 1646
Db       2003 ATGGGTGGCATGGGCGGCATGATGTA 2028


RESULT 9
US-08-470-260-7
; Sequence 7, Application US/08470260
; Patent No. 6077706
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    APPLICANT:  Bugnoli, Massimo
;    APPLICANT:  Telford, John
;    APPLICANT:  Macchia, Giovanni
;    APPLICANT:  Rappuoli, Rino
;    TITLE OF INVENTION:  Helicobacter Pylori Proteins Useful
```

; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1838 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-470-260-7

Query Match 39.0%; Score 642.4; DB 3; Length 1838;
Best Local Similarity 62.9%; Pred. No. 2.2e-159;
Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;





```
Db       1616 AAATCAAAGAAGAAAAAAGCGACTCCGGCAATGCCTGATATGGGTGGCATGGGCGGTATGG 1675

Qy       1625 GTGGTATGGACGGTATGTACTA 1646
Db       1676 GAGGCATGGGCGGCATGATGTA 1697
```

RESULT 10
US-08-471-491-7
; Sequence 7, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-08-471-491-7

Query Match 39.0%; Score 642.4; DB 3; Length 1838;
Best Local Similarity 62.9%; Pred. No. 2.2e-159;
Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;

```
Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db         62 CAAAAGAAATCAAATTTTCAGATAGTGCGAGAAACCTTTTATTTGAAGGCGTGAGGCAAC 121

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db        122 TCCATGACGCTGTCAAAGTAACCATGGGGCCAAGAGGCAGGAATGTATTGATCCAAAAAA 181

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db        182 GCTATGGCGCTCCAAGCATCACCAAAGACGGCGTGAGCGTGGCTAAAGAGATTGAATTAA 241

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db        242 GTTGCCCAGTAGCTAACATGGGCGCTCAACTCGTTAAAGAAGTAGCGAGCAAAACCGCTG 301

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db        302 ATGCTGCCGGCGATGGCACGACCACAGCGACCGTGCTAGCTTATAGCATTTTTAAAGAAG 361

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db        362 GTTTGAGGAATATCACGGCTGGGGCTAACCCTATTGAAGTGAAACGAGGCATGGATAAAG 421

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
Db        422 CTGCTGAAGCGATCATTAATGAGCTTAAAAAAGCGAGCAAAAAAGTAGGCGGTAAAGAAG 481

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
Db        482 AAATCACCCAAGTGGCGACCATTTCTGCAAACTCCGATCACAATATCGGGAAACTCATCG 541

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
Db        542 CTGACGCTATGGAAAAAGTGGGTAAAGACGGCGTGATCACCGTTGAGGAAGCTAAGGGCA 601

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
```

```
Db        602 TTGAAGATGAATTGGATGTCGTAGAAGGCATGCAATTTGATAGAGGCTACCTCTCCCCTT 661

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
Db        662 ATTTTGTAACGAACGCTGAGAAAATGACCGCTCAATTGGATAATGCTTACATCCTTTTAA 721

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
Db        722 CGGATAAAAAAATCTCTAGCATGAAAGACATTCTCCCGCTACTAGAAAAAACCATGAAAG 781

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
Db        782 AGGGCAAACCGCTTTTAATCATCGCTGAAGACATTGAGGGCGAAGCTTTAACGACTCTAG 841

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
Db        842 TGGTGAATAAATTAAGAGGCGTGTTGAATATCGCAGCGGTTAAAGCTCCAGGCTTTGGGG 901

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
Db        902 ACAGAAGAAAAGAAATGCTCAAAGACATCGCTATTTTAACCGGCGGTCAAGTCATTAGCG 961

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
Db        962 AAGAATTGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCTGGAAGGA 1021

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
Db       1022 TTGTGATTGACAAAGACAACACCACGATCGTAGATGGCAAAGGCCATAGCGATGATGTTA 1081

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
Db       1082 AAGACAGAGTCGCGCAGATCAAAACCCAAATTGCAAGTACGACAAGCGATTATGACAAAG 1141

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1142 AAAAATTGCAAGAAAGATTGGCTAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGCG 1201

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1202 CTGCGAGTGAAGTGGAAATGAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGCGCGA 1261

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1262 CTAAAGCGGCGGTTGAAGAAGGCATTGTGATTGGTGGCGGTGCGGCTCTCATTCGCGCGG 1321

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1322 CTCAAAAAGTGCATTTGAATTTGC------ACGATGATGAAAAAGTGGGCTATGAAATCA 1375

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1376 TCATGCGCGCCATTAAAGCCCCATTAGCTCAAATCGCTATCAACGCTGGTTATGATGGCG 1435

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1436 GTGTGGTCGTGAATGAAGTAGAAAAACACGAAGGGCATTTTGGTTTTAACGCTAGCAATG 1495

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1496 GCAAGTATGTGGATATGTTTAAAGAAGGCATTATTGACCCCTTAAAAGTAGAAAGGATCG 1555

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       1556 CTCTACAAAATGCGGTTTCGGTTTCAAGCCTGCTTTTAACCACAGAAGCCACCGTGCATG 1615

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTG---GCGGTGGTATGGGTGGTATGG 1624
Db       1616 AAATCAAAGAAGAAAAAGCGACTCCGGCAATGCCTGATATGGGTGGCATGGGCGGTATGG 1675

Qy       1625 GTGGTATGGACGGTATGTACTA 1646
Db       1676 GAGGCATGGGCGGCATGATGTA 1697
```

RESULT 11
US-08-466-662-7
; Sequence 7, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1838
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-08-466-662-7

Query Match 39.0%; Score 642.4; DB 3; Length 1838;
Best Local Similarity 62.9%; Pred. No. 2.2e-159;
Matches 1032; Conservative 0; Mismatches 601; Indels 9; Gaps 2;

```
Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
              |  |||||||||   |||   | ||  ||  |  |     | |   ||| ||     |||
Db         62 CAAAAGAAATCAAATTTTCAGATAGTGCGAGAAACCTTTTATTTGAAGGCGTGAGGCAAC 121

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
              |      | |||||  ||||||||   | ||  ||  | ||||  | |||||   |  ||   ||||
Db        122 TCCATGACGCTGTCAAAGTAACCATGGGGCCAAGAGGCAGGAATGTATTGATCCAAAAAA 181

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
                | |||  | |||   ||  ||  |||||  ||  ||    ||  ||  |||||  ||||||  |
Db        182 GCTATGGCGCTCCAAGCATCACCAAAGACGGCGTGAGCGTGGCTAAAGAGATTGAATTAA 241

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
                  |  ||  ||||||||||||||  | ||||||||||||||   ||||||     |
Db        242 GTTGCCCAGTAGCTAACATGGGCGCTCAACTCGTTAAAGAAGTAGCGAGCAAAACCGCTG 301

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
              ||  |||  ||  ||||  ||  ||  ||  ||  |     ||||  |    ||||
Db        302 ATGCTGCCGGCGATGGCACGACCACAGCGACCGTGCTAGCTTATAGCATTTTTAAAGAAG 361

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
              ||  |  |  |  |  |||||   |||  |||||  |  | |||||  |||||  |||||||
Db        362 GTTTGAGGAATATCACGGCTGGGGCTAACCCTATTGAAGTGAAACGAGGCATGGATAAAG 421

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
              |||  ||  ||  |  |  |||  ||  ||  |    |  ||         |  |  ||  |||
Db        422 CTGCTGAAGCGATCATTAATGAGCTTAAAAAAGCGAGCAAAAAAGTAGGCGGTAAAGAAG 481

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
              ||||  |  |||||  |  ||||||||||||||||  |||    |  ||  ||  ||||  |
Db        482 AAATCACCCAAGTGGCGACCATTTCTGCAAACTCCGATCACAATATCGGGAAACTCATCG 541

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
              ||||  |||||||  |||||  ||  ||||  ||  ||  |||||  ||||||||||||  ||
Db        542 CTGACGCTATGGAAAAAGTGGGTAAAGACGGCGTGATCACCGTTGAGGAAGCTAAGGGCA 601

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
              |||||  |  |||  |||||  ||  |||||  |||  |  |||||  |  ||||||||||  ||  |
Db        602 TTGAAGATGAATTGGATGTCGTAGAAGGCATGCAATTTGATAGAGGCTACCTCTCCCCTT 661

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
              |  ||||||||  ||  ||||||||||||      |  ||  |  |||||  ||||  |||||||  |
Db        662 ATTTTGTAACGAACGCTGAGAAAATGACCGCTCAATTGGATAATGCTTACATCCTTTTAA 721

Qy        668 ATGAGAAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
              ||  |||||  ||  ||||||||||||||||  ||  ||  |  |||||  ||       ||||
Db        722 CGGATAAAAAAAATCTCTAGCATGAAAGACATTCTCCCGCTACTAGAAAAAACCATGAAAG 781

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
                  |  ||  ||  |  ||  ||  |||||||||  |  ||  ||  |||||  |  |  ||  ||  |
Db        782 AGGGCAAACCGCTTTTAATCATCGCTGAAGACATTGAGGGCGAAGCTTTAACGACTCTAG 841

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
              |  ||  |||||  |  |  ||  |  |  |  |  |||  ||  ||||||||  ||  |||||  |
Db        842 TGGTGAATAAATTAAGAGGCGTGTTGAATATCGCAGCGGTTAAAGCTCCAGGCTTTGGGG 901

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
              |  |  |  |||  |||||  |||  ||  |||||  |  ||  ||  ||  ||  |
Db        902 ACAGAAGAAAAGAAATGCTCAAAGACATCGCTATTTTAACCGGCGGTCAAGTCATTAGCG 961

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
              ||||  ||  |  |  ||  |||||  |  ||  |  ||||||  |  ||||  |  |
Db        962 AAGAATTGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCTGGAAGGA 1021

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
              |  ||  ||||||||||||  ||  ||  ||  |||||  |||||      ||  |     ||  |||  |||
Db       1022 TTGTGATTGACAAAGACAACACCACGATCGTAGATGGCAAAGGCCATAGCGATGATGTTA 1081

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
              |||  ||||  |  ||  |  |||||||  |  ||  |  ||||||||  |  |
Db       1082 AAGACAGAGTCGCGCAGATCAAAACCCAAATTGCAAGTACGACAAGCGATTATGACAAAG 1141

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
              |||||  |  ||||||  |  |  ||  |||||  |||  ||  ||  |||||  ||  |  ||  ||  |
Db       1142 AAAAATTGCAAGAAAGATTGGCTAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGCG 1201

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
              ||||  |  ||||  ||||||||||||||||  ||  ||  ||  ||  ||  ||  |  |  ||  |
Db       1202 CTGCGAGTGAAGTGGAAATGAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGCGCGA 1261

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
              |  |  |||  ||||||||||||||||  |||||  ||||||  |||  |  |||  |  |||  |
Db       1262 CTAAAGCGGCGGTTGAAGAAGGCATTGTGATTGGTGGCGGTGCGGCTCTCATTCGCGCGG 1321

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
              |  ||     |  ||  |  |  |  |||||||||  |  ||  |  |  ||||
Db       1322 CTCAAAAAGTGCATTTGAATTTGC------ACGATGATGAAAAAGTGGGCTATGAAATCA 1375

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
              ||  ||  |  ||  |||  ||  |||  ||||||  |||  ||  |||||  |||||  ||
Db       1376 TCATGCGCGCCATTAAAGCCCCATTAGCTCAAATCGCTATCAACGCTGGTTATGATGGCG 1435

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
              |  |  ||  ||  |  ||||  |||  |||  ||||||  |||||  |||  |
Db       1436 GTGTGGTCGTGAATGAAGTAGAAAAACACGAAGGGCATTTTGGTTTTAACGCTAGCAATG 1495

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
              |  |  ||||  ||  |  ||||||  ||  |  |||||  ||  ||||||  |  |  ||  |
Db       1496 GCAAGTATGTGGATATGTTTAAAGAAGGCATTATTGACCCCTTAAAAGTAGAAAGGATCG 1555

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
              |  ||||||||||  |  ||  ||  |  |  |||  ||||  ||||||  |  |  |  ||
Db       1556 CTCTACAAAATGCGGTTTCGGTTTCAAGCCTGCTTTTAACCACAGAAGCCACCGTGCATG 1615

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTG---GCGGTGGTATGGGTGGTATGG 1624
              |||  ||||  |||  |  |  |  |||||||  |||||  |||||  |||||||
Db       1616 AAATCAAAGAAGAAAAAGCGACTCCGGCAATGCCTGATATGGGTGGCATGGGCGGTATGG 1675

Qy       1625 GTGGTATGGACGGTATGTACTA 1646
              |  ||  ||||  |||  |||   ||
Db       1676 GAGGCATGGGCGGCATGATGTA 1697
```

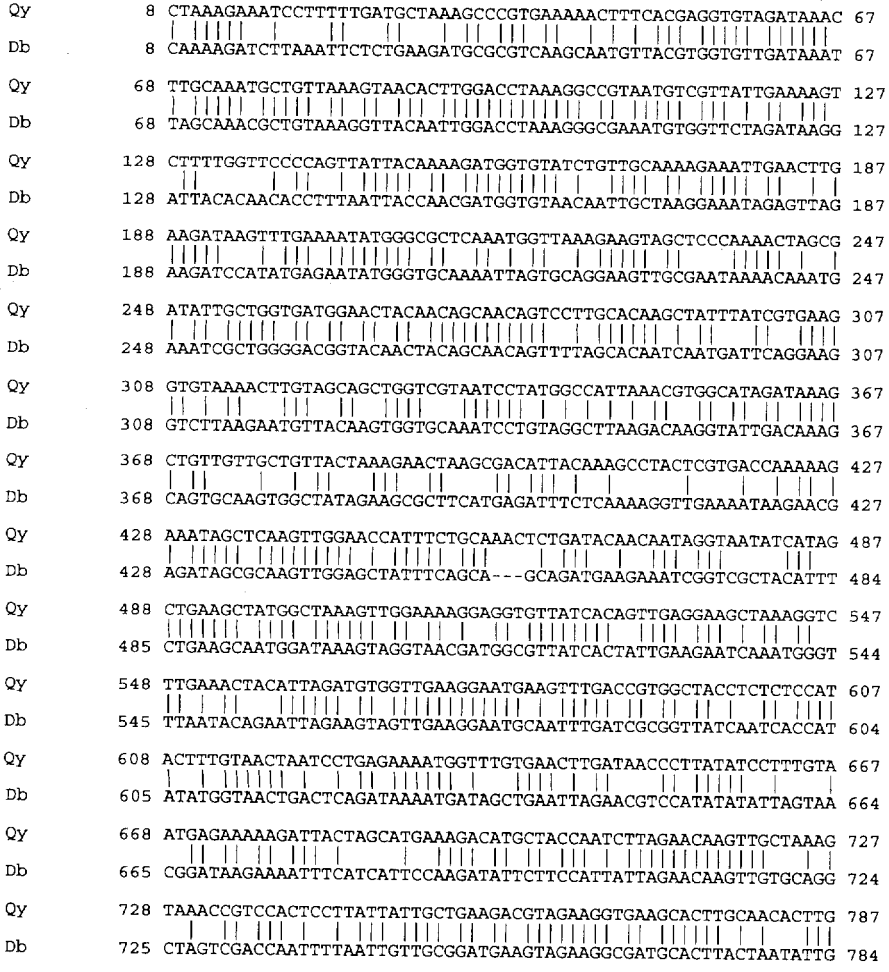
RESULT 12
US-08-256-847C-6
; Sequence 6, Application US/08256847C
; Patent No. 6403099

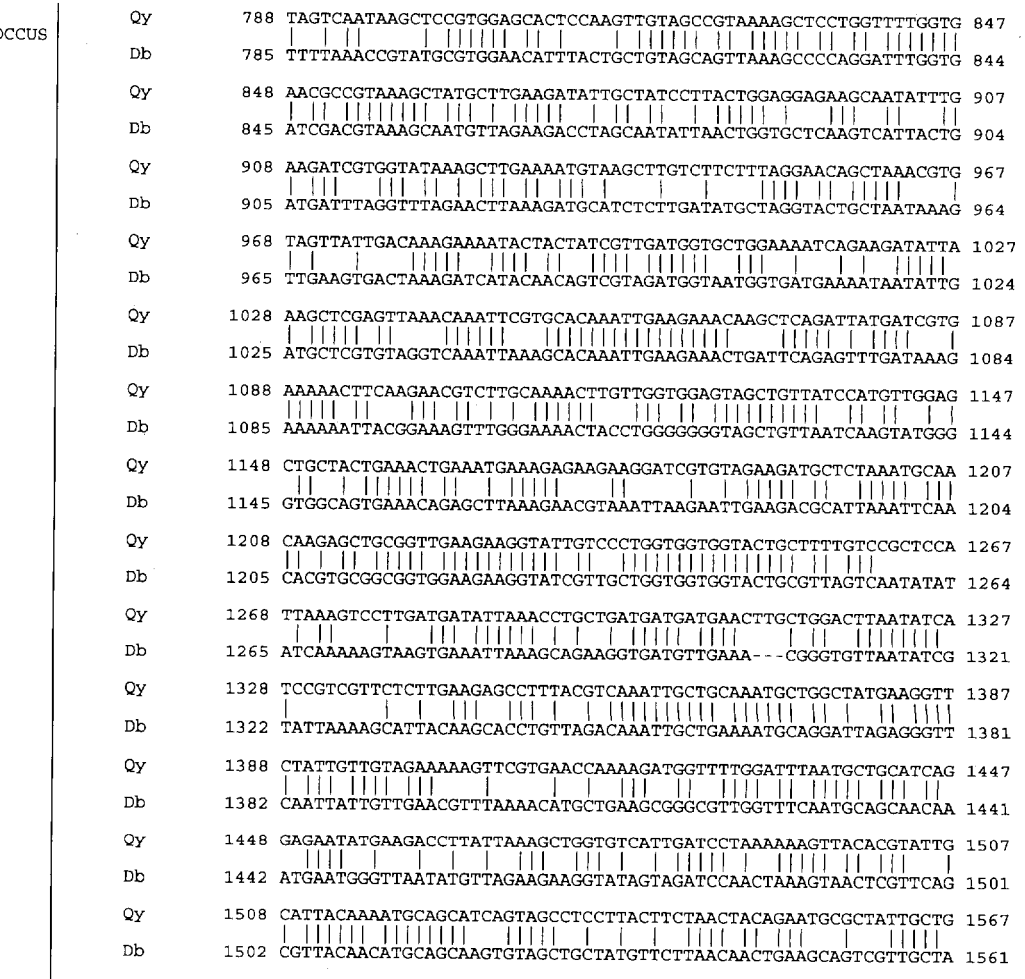
```
; GENERAL INFORMATION:
;  APPLICANT: Rappuoli, Rino
;  APPLICANT:  Costantino, Paolo
;  APPLICANT:  No. 6403099elli, Francesco
;  TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
;  FILE REFERENCE: CHIR-0042
;  CURRENT APPLICATION NUMBER: US/08/256,847C
;  CURRENT FILING DATE:  1994-11-01
;  PRIOR APPLICATION NUMBER: PCT/EP93/00516
;  PRIOR FILING DATE: 1993-03-08
;  PRIOR APPLICATION NUMBER: FI92A000058
;  PRIOR FILING DATE: 1992-03-06
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 1838
;   TYPE: DNA
;   ORGANISM: H. pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (58)..(1695)
;   OTHER INFORMATION:
US-08-256-847C-6

  Query Match            39.0%;  Score 642.4;  DB 4;  Length 1838;
  Best Local Similarity  62.9%;  Pred. No. 2.2e-159;
  Matches 1032;  Conservative    0;  Mismatches  601;  Indels    9;  Gaps    2;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
              | |||||||||   |||   | ||  ||  |  |     |  |   |||  ||     |||
Db         62 CAAAAGAAATCAAATTTTCAGATAGTGCGAGAAACCTTTTATTTGAAGGCGTGAGGCAAC 121

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
              |     | |||||  ||||||||  |  ||  ||  | ||||  | |||||  |  ||  ||||
Db        122 TCCATGACGCTGTCAAAGTAACCATGGGGCCAAGAGGCAGGAATGTATTGATCCAAAAAA 181

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
                | |||   | |||    ||  ||  |||||  ||  ||    ||  ||  |||||  ||||||  |
Db        182 GCTATGGCGCTCCAAGCATCACCAAAGACGGCGTGAGCGTGGCTAAAGAGATTGAATTAA 241

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
                  |  ||  ||||||||||||  | ||||||||||||||    ||||||     |
Db        242 GTTGCCCAGTAGCTAACATGGGCGCTCAACTCGTTAAAGAAGTAGCGAGCAAAACCGCTG 301

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
              ||  |||  ||  |||||  ||  ||  |||||  ||  ||  ||  ||    ||||  |   ||||
Db        302 ATGCTGCCGGCGATGGCACGACCACAGCGACCGTGCTAGCTTATAGCATTTTTAAAGAAG 361

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
              ||  |  |    |  |  |||||    |||  |||||  |    |  |||||  |||||  |||||||
Db        362 GTTTGAGGAATATCACGGCTGGGGCTAACCCTATTGAAGTGAAACGAGGCATGGATAAAG 421

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427
              |||  ||  ||  |  |  |||  ||  ||  |    |  ||          |  |  ||  |||
Db        422 CTGCTGAAGCGATCATTAATGAGCTTAAAAAAGCGAGCAAAAAAGTAGGCGGTAAAGAAG 481

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487
              ||||  |  |||||  |  ||||||||||||||||  |||   |  ||  ||  ||||  |
Db        482 AAATCACCCAAGTGGCGACCATTTCTGCAAACTCCGATCACAATATCGGGAAACTCATCG 541

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547
              ||||  |||||||  |||||  ||  ||||  ||  ||  |||||  ||||||||||||  ||
Db        542 CTGACGCTATGGAAAAAGTGGGTAAAGACGGCGTGATCACCGTTGAGGAAGCTAAGGGCA 601

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607
              |||||   |  |||  |||||  ||  |||||  |||  |  |||||   |  ||||||||||  ||  |
Db        602 TTGAAGATGAATTGGATGTCGTAGAAGGCATGCAATTTGATAGAGGCTACCTCTCCCCTT 661

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667
              |  ||||||||  ||  ||||||||||      |  ||  |  |||||   ||||  |||||||   |
Db        662 ATTTTGTAACGAACGCTGAGAAAATGACCGCTCAATTGGATAATGCTTACATCCTTTTAA 721

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727
```

```
                ||  |||||  ||  ||||||||||||||||  ||  ||   |  |||||  ||       ||||
Db        722 CGGATAAAAAAATCTCTAGCATGAAAGACATTCTCCCGCTACTAGAAAAAACCATGAAAG 781

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787
                  |   ||  ||  |  ||  ||  |||||||||  |  ||  ||  |||||  |  |  ||  | |
Db        782 AGGGCAAACCGCTTTTAATCATCGCTGAAGACATTGAGGGCGAAGCTTTAACGACTCTAG 841

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847
              |  ||  |||||  |  |  ||  |    |  |  |  |  |||  ||  ||||||||  ||  |||||  |
Db        842 TGGTGAATAAATTAAGAGGCGTGTTGAATATCGCAGCGGTTAAAGCTCCAGGCTTTGGGG 901

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907
              |  |  |  ||||  |||||  ||||  ||  |||||  |  ||  ||  |||  ||    |
Db        902 ACAGAAGAAAAGAAATGCTCAAAGACATCGCTATTTTAACCGGCGGTCAAGTCATTAGCG 961

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967
              ||||     ||  |  |  ||  |||||  |       ||    |  ||||||  |  ||||   |  |
Db        962 AAGAATTGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCTGGAAGGA 1021

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027
              |  ||  ||||||||||||  ||  ||  ||  |||||  |||||     ||   |    ||  |||  |||
Db       1022 TTGTGATTGACAAAGACAACACCACGATCGTAGATGGCAAAGGCCATAGCGATGATGTTA 1081

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087
              |||   ||||     ||  ||     |  |||||||  |    ||  |     ||||||||     |
Db       1082 AAGACAGAGTCGCGCAGATCAAAACCCAAATTGCAAGTACGACAAGCGATTATGACAAAG 1141

Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
              |||||  |  ||||||  |    |  ||  |||||     |||  ||  ||  |||||  ||    |  ||  ||  |
Db       1142 AAAAATTGCAAGAAAGATTGGCTAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGCG 1201

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
              ||||  |  ||||     ||||||||||||||  ||  ||  ||  ||  ||  ||  |  |  ||  |
Db       1202 CTGCGAGTGAAGTGGAAATGAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGCGCGA 1261

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
              |  |  |||  ||||||||||||||  |||||     ||||||  |||  |  |||  |  |  |||  |
Db       1262 CTAAAGCGGCGGTTGAAGAAGGCATTGTGATTGGTGGCGGTGCGGCTCTCATTCGCGCGG 1321

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
              |  ||     |  ||  |  |  |     |  ||||||||||  |  ||  |  ||||
Db       1322 CTCAAAAAGTGCATTTGAATTTGC------ACGATGATGAAAAAGTGGGCTATGAAATCA 1375

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
              ||  ||  |  ||  |||  ||  |||  ||||||  |||   ||  |||||  |||||  ||
Db       1376 TCATGCGCGCCATTAAAGCCCCATTAGCTCAAATCGCTATCAACGCTGGTTATGATGGCG 1435

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
              |  |  ||  ||  |  ||||     |||   |||  |||||  |||||  |||     |
Db       1436 GTGTGGTCGTGAATGAAGTAGAAAAACACGAAGGGCATTTTGGTTTTAACGCTAGCAATG 1495

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
              |  |  ||||  ||  |  ||||||  ||  |  ||||  ||   ||||||     |  |  ||  |
Db       1496 GCAAGTATGTGGATATGTTTAAAGAAGGCATTATTGACCCCTTAAAAGTAGAAAGGATCG 1555

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
              |  ||||||||||  |  ||  ||  |  |  |||  ||||  ||||||  |  |  |  ||
Db       1556 CTCTACAAAATGCGGTTTCGGTTTCAAGCCTGCTTTTAACCACAGAAGCCACCGTGCATG 1615

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTG---GCGGTGGTATGGGTGGTATGG 1624
              |||   ||||   |||      |  |  |||||||      |||||  |||||  |||||||
Db       1616 AAATCAAAGAAGAAAAAGCGACTCCGGCAATGCCTGATATGGGTGGCATGGGCGGTATGG 1675

Qy       1625 GTGGTATGGACGGTATGTACTA 1646
              |  ||  ||||  |||  |||    ||
Db       1676 GAGGCATGGGCGGCATGATGTA 1697

RESULT 13
US-09-134-001C-1868
; Sequence 1868, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
```

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1868
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1868

Query Match 38.9%; Score 640; DB 4; Length 1623;
Best Local Similarity 63.1%; Pred. No. 9e-159;
Matches 1020; Conservative 0; Mismatches 590; Indels 6; Gaps 2;





RESULT 14
US-09-221-017B-895/c
; Sequence 895, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  895:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2416 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...2416
US-09-221-017B-895

  Query Match             38.0%;  Score 626.4;  DB 4;  Length 2416;
  Best Local Similarity   62.0%;  Pred. No. 4.1e-155;
  Matches 1009;  Conservative    0;  Mismatches  616;  Indels    3;  Gaps    1;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67
Db       1924 CAAAAGAAATCAAATTCGATATGGAATCTCGCGACCTTCTGAAGAAGGGCGTAGATGCAC 1865
Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127
Db       1864 TGGCAAATGCCGTTAAAGTTACCCTCGGGCCGAAAGGTCGTAATGTTATCCTTAGCAAGA 1805
Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187
Db       1804 CGTACGGAGCTCCGCACATTACCAAGGACGGTGTGAGCGTAGCGAAAGAGATAGAATTGG 1745
Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247
Db       1744 AGTGCCCGTTCGAGAACATGGGTGCCCAGTTGGTGAAAGAAGTAGCCTCCAAGACCAATG 1685
Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307
Db       1684 ACGATGCCGGTGACGGTACGACTACGGCTACGATCCTCGCCCAGAGCATTATCGGCGTGG 1625
Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367
Db       1624 GTCTGAAGAACGTTACGGCAGGAGCCAATCCGATGGATTTGAAGCGTGGTATCGACAAGG 1565
Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAA---A 424
Db       1564 CTGTAAAGGCTGTGGTAACTCACATTGCAGGTATGGCTAAGGAAGTGGGCGACGACTTCC 1505
Qy        425 AAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCA 484
Db       1504 AGAAGATCGAGCACGTAGCCAAGATCTCTGCCAACGGAGACGAAAACATCGGTAGCCTCA 1445
Qy        485 TAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAG 544
Db       1444 TTGCCGAAGCCATGCGCAAGGTGAAGAAAGAAGGCGTTATCACGGTAGAAGAAGCCAAGG 1385
Qy        545 GTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTC 604
Db       1384 GAACCGACACTACGGTAGAAGTGGTGGAAGGTATGCAGTTCGACCGCGGCTACATCTCTC 1325
Qy        605 CATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTT 664
Db       1324 CCTACTTCGTAACGAACACGGATAAGATGGAGGTGCAAATGGAAAATCCTTTCATCCTCA 1265
Qy        665 GTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTA 724
Db       1264 TCTACGACAAGAAAATATCCGTCCTGAAAGAGATGCTCCCGATCCTCGAACAAACGGTTC 1205
Qy        725 AAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACAC 784
Db       1204 AGACGGGCAAGCCCCTCCTCATCATTGCAGAAGACATCGACAGCGAAGCACTCGCCACCT 1145
Qy        785 TTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTG 844
Db       1144 TGGTTGTAAACCGTCTGCGCGGCAGCCTCAAGATCTGTGCAGTGAAGGCTCCCGGATTCG 1085
Qy        845 GTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATAT 904
Db       1084 GCGATCGTCGCAAGGCTATGCTCGAAGACATTGCTATCCTGACGGGCGGAACGGTTATCA 1025
Qy        905 TTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAAC 964
Db       1024 GCGAAGAGACCGGTCTGAAGCTCGAAAATGCTACGATGGATATGCTCGGTACAGCTGAGA 965
Qy        965 GTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATA 1024
Db        964 AAGTAACGGTTGATAAGGACAATACTACTATCGTTAACGGAGCCGGAAACAAAGAAGGCA 905
Qy       1025 TTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATC 1084
Db        904 TCGCTTCACGTATCACGCAGATCAAAGCTCAGATCGAGAATACGACCAGCGACTACGACC 845
Qy       1085 GTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTG 1144
Db        844 GCGAAAAGCTGCAAGAACGTTTGGCCAAGCTCGCCGGCGGTGTAGCTGTTCTTTACGTGG 785
Qy       1145 GAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATG 1204
Db        784 GTGCTGCCAGCGAAGTGGAAATGAAGGAAAAGAAGGATCGCGTAGAAGATGCCTTGAGTG 725
Qy       1205 CAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCT 1264
Db        724 CAACGCGTGCTGCAATCGAAGAGGGTACAGTACCCGGTGGCGGTACGGCATACATTCGTG 665
Qy       1265 CCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATA 1324
Db        664 CCATAGCTGCTTTGGAAGGTCTCAAGGGTGAGAACGAAGACGAAACCCACAGGTATCGAGA 605
Qy       1325 TCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAG 1384
Db        604 TCGTGAAACGCGCCATCGAGGAGCCGCTTCGTCAGATCGTAGCGAACGCCGGTAAAGAGG 545
Qy       1385 GTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCAT 1444
Db        544 GTGCCGTTGTGGTACAGAAGGTGAAAGAAGGCAAGGACGACTTCGGCTACAATGCCCGTA 485
Qy       1445 CAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTA 1504
```

```
Db        484 CGGATGTTTTCGAAAACCTCTACACTACCGGTGTTATCGACCCGGCCAAAGTAACACGTG 425

Qy       1505 TTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTG 1564

Db        424 TAGCATTGGAAAATGCAGCGTCTATCGCAGGTATGTTCCTGACTACGGAGTGCGTTATCG 365

Qy       1565 CTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGG 1624

Db        364 CTGACAAGAAAGAAGATAATCCTGCCGCACCGGCCATGCCCGGAGGTATGGGAGGAATGG 305

Qy       1625 GTGGTATG 1632

Db        304 GCGGCATG 297


RESULT 15
US-08-467-822-28
; Sequence 28, Application US/08467822
; Patent No. 5843460
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,822
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-467-822-28

  Query Match           37.1%;  Score 610.4;  DB 2;  Length 2284;
  Best Local Similarity  62.1%;  Pred. No. 6.5e-151;
  Matches 1019;  Conservative    0;  Mismatches  611;  Indels   12;  Gaps    3;

Qy          8 CTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAAC 67

Db        510 CAAAAGAAATCAAATTTTCAGATAGCGCAAGAAACCTTTTATTTGAAGGCGTAAGACAAC 569

Qy         68 TTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAGT 127

Db        570 TCCATGACGCTGTCAAAGTAACCATGGGGCCAAGAGGCAGGAACGTGTTGATCCAAAAAA 629

Qy        128 CTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTTG 187

Db        630 GCTATGGCGCTCCAAGCATCACCAAAGACGGCGTGAGCGTGGCTAAAGAGATTGAATTAA 689

Qy        188 AAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGCG 247

Db        690 GTTGCCCCGTGGCTAACATGGGCGCTCAGCTCGTTAAAGAAGATGCGAGCAAAACCGCTG 749

Qy        248 ATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAAG 307

Db        750 ATGCCGCCGGCGATGGCACGACCACAGCGACCGTGCTGGCTTATAGCATTTTTAAAGAGG 809

Qy        308 GTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAAG 367

Db        810 GCTTGAGGAATATCACGGCTGGGGCTAACCCTATTGAAGTGAAACGAGGCATGGATAAAG 869

Qy        368 CTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAAG 427

Db        870 CGCCTGAAGCGATCATTAATGAGCTTAAAAAAGCGAGCAAAAAAGTGGGCGGTAAAGAAG 929

Qy        428 AAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATAG 487

Db        930 AAATCACCCAAGTAGCGACCATTTCTGCAAACTCCGATCACAATATCGGGAAACTCATCG 989

Qy        488 CTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGTC 547

Db        990 CTGACGCTATGGAAAAAGTGGGTAAAGACGGCGTGATCACCGTTGAAGAAGCTAAGGGCA 1049

Qy        548 TTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCAT 607

Db       1050 TTGAAGATGAATTAGATGTCGTAGAAGGCATGCAATTTGATAGAGGCTACCTCTCCCCTT 1109

Qy        608 ACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGTA 667

Db       1110 ACTTTGTAACCAACGCTGAGAAAATGACCGCTCAATTGGATAACGCTTACATCCTTTTAA 1169

Qy        668 ATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAAG 727

Db       1170 CGGATAAAAAAATCTCTAGCATGAAAGACATTCTCCCGCTACTAGAAAAAACCATGAAAG 1229

Qy        728 TAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTTG 787

Db       1230 AGGGCAAACCGCTTTTAATCATCGCTGAAGACATTGAGGGCGAAGCTTTAACGACTCTAG 1289

Qy        788 TAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGTG 847

Db       1290 TGGTGAATAAATTAAGAGGCGTGTTGAATATCGCAGCGGTTAAAGCTCCAGGCTTTGGGG 1349

Qy        848 AACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTTG 907

Db       1350 ACAGGAGAAAAGAAATGCTCAAAGACATCGCTGTTTTAACCGGCGGTCAAGTCATTAGCG 1409

Qy        908 AAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGTG 967

Db       1410 AAGAATTGGGCTTGAGTCTAGAAAACGCTGAAGTGGAGTTTTTAGGCAAAGCGAA---GA 1466

Qy        968 TAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATTA 1027

Db       1467 TTGTGATTGACAAAGACAACACCACGATCGTAGATGGCAAAGGCCATAGCCATGACGTCA 1526

Qy       1028 AAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGTG 1087

Db       1527 AAGACAGAGTCGCGCAAATCAAAACCCAAATTGCAAGCACGACAAGCGATTACGACAAAG 1586
```

```
Qy       1088 AAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGAG 1147
Db       1587 AAAAATTGCAAGAAAGATTGGCCAAACTCTCTGGCGGTGTGGCTGTGATTAAAGTGGGCG 1646

Qy       1148 CTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCAA 1207
Db       1647 CTGCGAGTGAAGTGGAAATGAAAGAGAAAAAAGACCGGGTGGATGACGCGTTGAGCGCGA 1706

Qy       1208 CAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCCA 1267
Db       1707 CTAAAGCGGCGGTTGAAGAAGGCATTGTGATTGGGGGCGGTGCGGCCCTCATTCGCGCGG 1766

Qy       1268 TTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATCA 1327
Db       1767 CCCAAAAAGTGCATTTGAATTTAC------ACGATGATGAAAAAGTGGGCTATGAAATCA 1820

Qy       1328 TCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGTT 1387
Db       1821 TCATGCGCGCCATTAAAGCCCCATTAGCTCAAATCGCTATCAATGCCGGTTATGATGGCG 1880

Qy       1388 CTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCAG 1447
Db       1881 GTGTGGTCGTGAATGAAGTAGAAAAACACGAAGGGCATTTTGGTTTTAACGCTAGCAATG 1940

Qy       1448 GAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATTG 1507
Db       1941 GCAAGTATGTGGACATGTTTAAAGAAGGCATTATTGACCCCTTAAAAGTAGAAAGGATCG 2000

Qy       1508 CATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCTG 1567
Db       2001 CTTTACAAAATGCGGTTTCGGTTTCAAGCCTGCTTTTAACCACAGAAGCCACCGTGCATG 2060

Qy       1568 AAAAACCAGAACCTAAAAAAGATATGCCTATGCCTG---GCGGTGGTATGGGTGGTATGG 1624
Db       2061 AAATCAAAGAAGAAAAAGCGGCCCCAGCAATGCCTGATATGGGTGGCATGGGCGGAATGG 2120

Qy       1625 GTGGTATGGACGGTATGTACTA 1646
Db       2121 GAGGCATGGGCGGCATGATGTA 2142
```

Search completed: January 29, 2004, 15:27:33
Job time : 105 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:18:26 ; Search time 18.5 Seconds
(without alignments)
8373.295 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125846_11606/app_query.fasta_1.1799
-DB=SwissProt_41 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574_@CGN_1_1_18_@runat_28012004_125846_11606 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	1891.5	63.7	547	1	CH60_RALSO	Q8y1p8	ralstonia s
2	1880	63.3	546	1	CH60_ACEAC	Q8gbd2	acetobacter
3	1872.5	63.1	546	1	CH60_BURCE	Q9zfe0	burkholderi
4	1868.5	63.0	546	1	CH60_BURPS	Q9f712	burkholderi
5	1868.5	63.0	546	1	CH60_BURVI	Q9zfd8	burkholderi
6	1868	62.9	546	1	CH60_RHOPA	Q93mh1	rhodopseudo
7	1867.5	62.9	546	1	CH60_BURTH	P58723	burkholderi
8	1867.5	62.9	547	1	CH60_BARHE	O33963	bartonella
9	1867	62.9	545	1	CH61_RHIME	P35469	rhizobium m
10	1866	62.9	551	1	CH60_AMOPS	P26004	amoeba prot
11	1864.5	62.8	547	1	CH60_BARQU	O33964	bartonella
12	1864	62.8	545	1	CH64_RHIME	Q92zq4	rhizobium m
13	1862.5	62.8	547	1	CH60_XYLFA	Q9pfp2	xylella fas
14	1852.5	62.4	540	1	CH60_RHOMR	Q9xca9	rhodothermu
15	1852.5	62.4	546	1	CH60_CHRVI	P31293	chromatium
16	1847	62.2	546	1	CH60_XANAC	Q8ppz1	xanthomonas
17	1845	62.2	545	1	CH60_RHOCA	P95678	rhodobacter
18	1844.5	62.1	546	1	CH60_RHILV	P34939	rhizobium l
19	1844	62.1	546	1	CH60_XANCH	Q8rit7	xanthomonas
20	1843.5	62.1	547	1	CH60_BORPE	P48210	bordetella
21	1843.5	62.1	552	1	CH60_COXBU	P19421	coxiella bu
22	1842	62.1	546	1	CH60_XANCP	Q8pd23	xanthomonas
23	1841	62.0	542	1	CH65_RHIME	P35471	rhizobium m
24	1840.5	62.0	544	1	CH60_AGRT5	P30779	agrobacteri
25	1834.5	61.8	546	1	CH61_RHOSH	P20110	rhodobacter
26	1834.5	61.8	552	1	CH60_PSEST	O33500	pseudomonas
27	1834	61.8	542	1	CH62_RHILE	Q9l691	rhizobium l
28	1833.5	61.8	539	1	CH61_BRAJA	P77829	bradyrhizob
29	1832.5	61.7	545	1	CH60_PARDE	Q9z462	paracoccus
30	1828.5	61.6	545	1	CH60_CHLTE	Q8kf02	chlorobium
31	1826	61.5	542	1	CH62_RHILO	Q98ih9	rhizobium l
32	1825.5	61.5	547	1	CH60_PSEAE	P30718	pseudomonas
33	1823	61.4	546	1	CH60_BRUAB	P25967	brucella ab
34	1819	61.3	546	1	CH60_BRUSU	Q8fx87	brucella su
35	1818	61.3	549	1	CH62_BRAJA	P35861	bradyrhizob
36	1816	61.2	542	1	CH62_RHIME	P35470	rhizobium m
37	1814.5	61.1	544	1	CH60_NEIMA	P57006	neisseria m
38	1814	61.1	546	1	CH60_BRUME	Q8yb53	brucella me
39	1812.5	61.1	551	1	CH64_RHILO	Q983s4	rhizobium l
40	1810	61.0	552	1	CH63_RHILO	Q98ax9	rhizobium l
41	1808.5	60.9	544	1	CH60_NEIMB	P42385	neisseria m
42	1805.5	60.8	543	1	CH60_BARBA	P35635	bartonella
43	1804.5	60.8	547	1	CH60_ALTHA	Q9xau7	alteromonas
44	1803.5	60.8	549	1	CH65_RHILO	Q981j9	rhizobium l
45	1802.5	60.7	544	1	CH60_NEIGO	P29842	neisseria g

ALIGNMENTS

RESULT 1
CH60_RALSO
ID CH60_RALSO STANDARD; PRT; 547 AA.
AC Q8Y1P8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL OR MOPA OR RSC0642 OR RS01546.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC --

CC --
DR EMBL; AL646060; CAD14172.1; -.

```
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Complete proteome.
SQ   SEQUENCE   547 AA;  57404 MW;  0487E4650867C25A CRC64;

Alignment Scores:
Pred. No.:              2.32e-97      Length:         547
Score:                  1891.50       Matches:        380
Percent Similarity:     82.27%        Conservative:   70
Best Local Similarity:  69.47%        Mismatches:     96
Query Match:            63.73%        Indels:         1
DB:                     1             Gaps:           1

US-09-077-574A-1 (1-1647) x CH60_RALSO (1-547)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::::|||      ||||||   |||:::      ||||||
Db          1 MetAlaAlaLysAspValValPheGlyAspAlaAlaArgAlaLysMetValGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              |||:::||||||||    |||   :::|||||||||||||||||||||||||||||||
Db         41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::||||||   :::|||||||||||||||||||||||||||||||   |||
Db         61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||||||||   ||||||||||||||||||||||||||||||||||:::|||
Db         81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||:::|||   ||||||||||||   ||||||||   :::||||||||||||
Db        101 ArgGluGlyMetLysTyrValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||      :::||||||      |||:::|||||||||
Db        121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProThrThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||||   ||||||||||||||||||   :::||||||
Db        141 SerLysGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspGluSerIleGlyAla 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||   ||||||||||||   ||||||||||||||||||:::
Db        161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||   ||||||      ||||||||||||||||||||:::||||||||||||||||||
Db        181 LysSerLeuGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||:::   ||||||||||   |||   :::||||||||||::::::
Db        201 SerProTyrPheIleAsnAsnProGluLysGlnValValGlnLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||      :::|||||||||:::::::::::|||:::||||||:::||||||||||||
Db        221 LeuLeuPheAspLysLysIleSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      ||||||||||||||:::|||||||||||||||||||||||||||||
Db        241 AlaLysAlaGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   |||:::      ||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::|||:::   ||||||   |||
Db        301 IleAlaGluGluValGlyLeuThrLeuGluLysAlaThrLeuAsnAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||||||||   |||   |||||||||||||||||:::||||||||||||   :::
Db        321 LysArgValGluIleGlyLysGluAsnThrThrIleIleAspGlyAlaGlyAspAlaArg 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              :::|||:::|||||||||||||||:::|||||||||||||||||   :::||||||||||
Db        341 AsnIleGluAlaArgValLysGlnValArgAlaGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||||:::|||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              |||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||||||||||   |||||||||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::      ::::::      :::|||   |||:::   |||:::   ||||||:::
Db        421 ArgAlaArgAlaLeuIleSerGlyLeuLysGlyAlaAsnAlaAspGlnAspAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                |||:::   |||::::::|||||||||||||||||||||      |||||||||
Db        441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnAlaGlyAsp 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::|||||      |||          |||      :::|||:::||||||
Db        461 GluAlaSerValValValAlaAsnValIleAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ::::::|||||||||   |||||::::::   |||||:::||||||   |||||||||
Db        481 SerThrGlyGluTyrGlyAspLeuValGluMetGlyValLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   ||||||||||||||||||||||||||||||:::|||||||||:::||||||
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaSerLeuMetLeuThrThrAspCysAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              :::||||||   |||:::                  ||||||   |||||||||||||||
Db        521 ValAlaGluLeuProLysAspAspAlaAlaProAlaMetPro---GlyGlyMetGlyGly 539

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
              |||||||||||||||||||||
Db        540 MetGlyGlyMetAspGlyMet 546

RESULT 2
CH60_ACEAC
ID   CH60_ACEAC     STANDARD;      PRT;   546 AA.
AC   Q8GBD2;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Acetobacter aceti.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Acetobacter.
OX   NCBI_TaxID=435;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 3283;
RA   Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
RA   Yanagida F.;
RT   "Cloning and Characterization of groEL Operon in Acetobacter aceti.";
RL   J. Biosci. Bioeng. 94:140-147(2002).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB081586; BAC16232.1; -.
DR   HAMAP; MF_00600; -; 1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  58106 MW;  3E476AF42F2CD5CC CRC64;
```

Alignment Scores:

Pred. No.:	9.99e-97	Length:	546
Score:	1880.00	Matches:	381
Percent Similarity:	82.42%	Conservative:	69
Best Local Similarity:	69.78%	Mismatches:	94
Query Match:	63.34%	Indels:	2
DB:	1	Gaps:	2

US-09-077-574A-1 (1-1647) x CH60_ACEAC (1-546)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::   |||  |||  ||||||:::::::      ||||||
Db          1 MetAlaAlaLysAspValLysPheGlyAlaAspAlaArgGlnArgMetArgAlaGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||  ||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||||:::|||  ||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||  |||||||||||||||||||||||||||||:::::::||||||||   |||
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||  ||||||||||||  |||||||||  :::||||||||||||
Db        101 ArgGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||       |||  :::||||||              |||
Db        121 AspLysAlaValAlaValValIleGluGluLeuLysLysAsnAlaLysLysValThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                  |||  |||||||||||||||||||||||||||  ::::::   ||||||
Db        141 ProAlaGluThrAlaGlnValGlyThrIleSerAlaAsnGlyGluSerGluIleGlyGln 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::|||:::|||||||||   |||||||||      ||||||||||||||||||||||||
```

```
Db        161 MetIleSerGluAlaMetGlnLysValGlySerGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      :::|||   |||||||||||||||||||:::|||||||||||||||||:::
Db        181 LysHisPheGlnThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrIle 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||      :::|||:::||||||||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetThrAlaAspLeuGluAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||:::::::|||::::::   |||||||||:::||||||   |||
Db        221 LeuIleHisGluLysLysLeuSerSerLeuGlnProMetLeuProLeuLeuGluSerVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::      |||||||||||||||||||||||||||||:::|||||||||||||||||
Db        241 ValGlnSerGlyArgProLeuLeuIleIleAlaGluAspValAspGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||   |||::::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||      ||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgAsnValMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||  ||||||  ||||||||||||||||  |||:::|||:::  ||||||||||||
Db        301 IleSerGluAspLeuGlyIleLysLeuGluThrValThrLeuAsnMetLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   ||||||||||||||||||||||||||||||||||||||||||:::::
Db        321 LysLysValHisIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysAlaAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspIleLysGlyArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   :::||||||   |||:::||||||:::|||||||||||:::||||||||||
Db        381 ValGlyGlySerThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaLeuAla 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::         |||:::   :::         :::||||||:::      |||
Db        421 ArgAlaThrLeuLysLeuGluGlyLeuHisTyrHisAsnAspAspGlnArgValGlyGly 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              :::||||||||||||:::|||:::  ||||||||||||||||  |||||||||
Db        441 AspIleIleArgArgAlaLeuGlnAlaProLeuArgGlnIleAlaHisAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
              :::|||:::::::::      ||||||  |||      |||      ||||||||:::
Db        461 AspGlyAlaValIleAlaAsnLysValLeuGluAsnSerAspTyrAsnPheGlyPheAsp 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
              |||   :::||||||||:::::|||:::::|||||:::|||||||||   ||||||
Db        481 AlaGlnAlaGlyGluTyrLysAsnLeuValGluAlaGlyIleIleAspProAlaLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
              |||  ||||||||:::|||||||||||||||   ||||||:::|||||||||
Db        501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuIleThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
              :::||||:::||||||   ||||||   |||      |||      ||||||
Db        521 MetValAlaGluArgProGlu---LysLysAlaAlaProAlaGlyGlyProAspMetGly 539
```

```
Qy       1618 GGTATGGGTGGTATGGAC 1635
              ||||||||||||||||||
Db        540 GlyMetGlyGlyMetAsp 545

RESULT 3
CH60_BURCE
ID   CH60_BURCE     STANDARD;      PRT;   546 AA.
AC   Q9ZFE0;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Burkholderia cepacia (Pseudomonas cepacia).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 10744;
RA   Zysk G., Splettstoesser W.D., Neubauer H.;
RT   "Nucleotide sequence comparison of the groE operon of Burkholderia
RT   spp.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF104907; AAC79087.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  56980 MW;  174B9934345E7315 CRC64;

Alignment Scores:
Pred. No.:                2.59e-96      Length:          546
Score:                    1872.50       Matches:         374
Percent Similarity:       82.08%        Conservative:    75
Best Local Similarity:    68.37%        Mismatches:      95
Query Match:              63.09%        Indels:          3
DB:                       1             Gaps:            2

US-09-077-574A-1 (1-1647) x CH60_BURCE (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::::|||         ||||||   |||:::      ||||||
Db          1 MetAlaAlaLysAspValValPheGlyAspSerAlaArgSerLysMetValGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::  ||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              |||:::|||||||||   |||  :::|||||||||||||||||||||||||||||||||
```

```
Db         41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::|||||||   :::|||||||||||||||||||||||||||||||   |||
Db         61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||||||||   ||||||||||||||||||||||||||||||||||||:::|||
Db         81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||||:::|||   |||||:::|||   |||||||||   :::|||||||||||||
Db        101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||   :::||||||   |||:::||||||
Db        121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||||||||||||||||||:::|||||||||||||||||||:::::|||||||:::
Db        141 AsnLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspSerSerIleGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||   ||||||||||||   |||||||||||||||||||:::
Db        161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||   |||   ||||||||||||||||||||:::|||||||||||||||||||
Db        181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||:::   ||||||:::|||   |||   ||||||||||||:::::
Db        201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   ::::::||||||:::::::::::::::|||:::||||||:::|||||||||||
Db        221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysAlaGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   |||:::   ||||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::|||:::   ||||||   |||
Db        301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||:::   :::   ||||||||||||||||||:::|||||||||||||:::::::
Db        321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              |||:::|||||||||||||||:::|||||||||||||||||||   :::|||||||||
Db        341 SerIleGluAlaArgValLysGlnValArgAlaGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||:::|||||||||   ||||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||   ||||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||   |||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIle 420
```

```
Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaArgThrAlaIleAlaGlyLeuThrGlyAlaAsnAlaAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluAlaSerValValValAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaThrGlyGluTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        521 ValAlaGluLeu------ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
Db        539 MetGlyGlyMet---GlyMet 544
```

```
RESULT 4
CH60_BURPS
ID   CH60_BURPS     STANDARD;      PRT;   546 AA.
AC   Q9F712;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=28450;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Woo P.C.Y., Leung P.K.L.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF287633; AAG32927.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57145 MW;  E9E9366EE8BD6ABD CRC64;
```

```
Alignment Scores:
Pred. No.:                4.3e-96        Length:          546
Score:                    1868.50        Matches:         373
Percent Similarity:       81.90%         Conservative:    75
Best Local Similarity:    68.19%         Mismatches:      96
Query Match:              62.95%         Indels:          3
DB:                       1              Gaps:            2

US-09-077-574A-1 (1-1647) x CH60_BURPS (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValValPheGlyAspSerAlaArgAlaLysMetValGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 AsnLysGluIleAlaGlnValGlyAlaIleSerAlaAsnSerAspSerSerIleGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuGluAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
```

```
Db        301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||:::   :::   |||||||||||||||||||:::|||||||||||:::::
Db        321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaVal 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              :::|||:::|||||||||||||||||||||   ||||||||||   :::||||||||
Db        341 AsnIleGluAlaArgValLysGlnIleArgThrGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||:::|||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||||||||||||||||||   |||   :::
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::        :::     :::          :::   |||:::   ||||||:::
Db        421 ArgAlaArgThrAlaIleAlaSerLeuThrGlyValAsnAlaAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 |||:::   |||::::::|||||||||||||||||||||||||      |||   |||
Db        441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::||||||      |||          |||      :::|||:::||||||
Db        461 GluAlaSerValValValAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::|||||||||   |||:::::::::|||||||||:::||||||   |||||||||
Db        481 AlaThrGlyGluTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   |||||||||||||||||||||||||   |||||||||||||:::   |||
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              :::||||||          ||||||:::|||   |||||||||||||||   ||||||
Db        521 ValAlaGluLeu------ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
              ||||||||||||   ||||||
Db        539 MetGlyGlyMet---GlyMet 544

RESULT 5
CH60_BURVI
ID   CH60_BURVI     STANDARD;      PRT;   546 AA.
AC   Q9ZFD8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Burkholderia vietnamiensis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=60552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 11319;
RA   Zysk G., Splettstoesser W.D., Neubauer H.;
RT   "Nucleotide sequence comparison of the groE operon of Burkholderia
RT   spp.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF104908; AAC79089.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57010 MW;  47FA893959221218 CRC64;

Alignment Scores:
Pred. No.:                4.3e-96        Length:          546
Score:                    1868.50        Matches:         373
Percent Similarity:       81.90%         Conservative:    75
Best Local Similarity:    68.19%         Mismatches:      96
Query Match:              62.95%         Indels:          3
DB:                       1              Gaps:            2

US-09-077-574A-1 (1-1647) x CH60_BURVI (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::::|||      ||||||   |||:::      ||||||
Db          1 MetAlaAlaLysAspValValPheGlyAspSerAlaArgSerLysMetValGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   |||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              |||:::|||||||||   |||   :::|||||||||||||||||||||||||||||||||
Db         41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||:::||||||   :::|||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||||||||   |||||||||||||||||||||||||||||||||||||||:::|||
Db         81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||:::|||   ||||||:::|||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||   |||      :::||||||      |||:::||||||
Db        121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||||||||||||||||||:::|||||||||||||||||:::::|||||:::
Db        141 AsnLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspSerSerIleGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||   ||||||||||||   ||||||||||||||||||:::
Db        161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||   |||          |||||||||||||||||||:::|||||||||||||||||||
```

```
Db        181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::   ||||||:::|||   |||      ||||||||||||:::::: 
Db        201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuAspAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   ::::::|||||:::::::::::::::|||:::||||||:::|||||||||||| 
Db        221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||      ||||||||||||||||||||||||||||||||||||||||||||||| 
Db        241 AlaLysAlaGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   ::||||||   |||:::   |||||||||||||||||||| 
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||:::  
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::|||:::   ||||||   ||| 
Db        301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||:::   :::   |||||||||||||||||:::||||||||||||:::::: 
Db        321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                 |||:::|||||||||||:::|||   |||||||||||   :::|||||||| 
Db        341 SerIleGluAlaArgValLysGlnValArgThrGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||:::|||||||||   |||||||||||||||||| 
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||   ||||||||||||||||   |||||||||||||||| 
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||   |||||||||   |||   ::: 
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::         :::      :::      |||:::   |||:::   ||||||::: 
Db        421 ArgAlaArgThrAlaIleAlaGlyLeuThrGlyAlaAsnAlaAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                 |||:::   |||::::::|||||||||||||||||||||||      |||   ||| 
Db        441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              |||   |||:::||||||      |||          |||      :::|||:::||||||  
Db        461 GluAlaSerValValValAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              |||:::|||||||||   |||:::::::::|||||||:::||||||   |||||||| 
Db        481 AlaThrGlyGluTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              |||   |||||||||||||||||||||||||   |||||||||||||:::   ||| 
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              :::||||||          ||||||:::|||   ||||||||||||||      |||||| 
Db        521 ValAlaGluLeu------ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
              ||||||||||||   |||||| 
Db        539 MetGlyGlyMet---GlyMet 544
```

```
RESULT 6
CH60_RHOPA
ID   CH60_RHOPA     STANDARD;      PRT;   546 AA.
AC   Q93MH1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Rhodopseudomonas palustris.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Rhodopseudomonas.
OX   NCBI_TaxID=1076;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xiao M., Zhu C.R., Qian X.M., Zheng P., Chen Y.Y.;
RT   "Cloning and Sequencing of the groESL Operon of Rhodopseudomonas
RT   palustris.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF406639; AAK94943.1; -.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57937 MW;  163C7115D68FFFF8 CRC64;
```

Alignment Scores:			
Pred. No.:	4.59e-96	Length:	546
Score:	1868.00	Matches:	374
Percent Similarity:	81.93%	Conservative:	75
Best Local Similarity:	68.25%	Mismatches:	95
Query Match:	62.94%	Indels:	4
DB:	1	Gaps:	2

US-09-077-574A-1 (1-1647) x CH60_RHOPA (1-546)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||::::::   ||||||      |||||:::::::::   ||||||||| 
Db          1 MetAlaAlaLysAspValLysPheAspThrAspAlaArgAspArgMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                 :::   ||||||:::||||||||||||||||||||||||||||||||||||||||| 
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
                 :::|||||||||||||:::|||   |||||||||||||||||||||||||||||||| 
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||   ::: 
Db         61 GluLeuSerAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerArg 80
```

```
Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspGluAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 LysGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaThrAlaLysValValGluSerIleLysAlaAlaSerArgProValAsnAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 GlnHisGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluAlaGlnIleGlyArg 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 PheIleAlaAspAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAsn 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysGlyLeuGluThrGluValGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetThrAlaGluLeuGluAspValPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerSerLeuGlnProMetValProLeuLeuGluSerVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 IleGlnAlaGlnArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyMetLysLeuGluAsnValThrIleAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleAsnLysAspAsnThrThrIleValAspGlyAlaGlyGluLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleGluAlaArgValSerGlnIleArgThrGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyMetThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValValGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 GlnAlaGlyLysValLeuAspGlyLeuThrGlyGluAsnProAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 ThrIleValArgArgAlaLeuGluAlaProLeuArgGlnIleAlaGlnAsnAlaGlyVal 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
Db        461 AspGlySerValValAlaGlyLysValArgGluSerAspAspLysAlaPheGlyPheAsn 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        481 AlaGlnThrGluGluTyrGlyAspMetPheLysPheGlyValIleAspProAlaLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        501 ValArgThrAlaLeuGluAspAlaAlaSerValAlaSerLeuLeuIleThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db        521 MetIleAlaAspLysProGluProLysSer---------AlaProAlaGlyGlyMetGly 537

Qy       1618 GGTATGGGTGGTATGGACGGTATG 1641
Db        538 GlyMetGlyGlyMetAspGlyMet 545
```

RESULT 7
CH60_BURTH
ID CH60_BURTH STANDARD; PRT; 546 AA.
AC P58723;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL.
OS Burkholderia thailandensis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=57975;
RN [1]
RP SEQUENCE FROM N.A.
RA Woo P.C.Y., Woo G.K.S., Yuen K.Y.;
RT "Burkholderia thailandensis chaperonin GroEL gene, complete
RT sequence.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC   proper assembly of unfolded polypeptides generated under stress
CC   conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC   7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF454383; AAL49762.1; -.
DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 57145 MW; 913CD01A5C131A06 CRC64;

Alignment Scores:

Pred. No.:	4.89e-96	Length:	546
Score:	1867.50	Matches:	373
Percent Similarity:	82.08%	Conservative:	76

```
Best Local Similarity:  68.19%        Mismatches:    95
Query Match:            62.92%        Indels:         3
DB:                     1             Gaps:           2

US-09-077-574A-1 (1-1647) x CH60_BURTH (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValValPheGlyAspSerAlaArgAlaLysMetValGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluArgSerPheGlyGlyProThrValThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysLeuGlnAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrSerAspAsnAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyMetLysTyrValAlaSerGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValAlaAlaAlaValGluGluLeuLysLysIleSerLysProCysThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 AsnLysGluIleAlaGlnValGlySerIleSerAlaAsnSerAspSerSerIleGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 ArgIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysSerLeuAlaAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnAsnProAspLysGlnValAlaValLeuGluAsnProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisAspLysLysValSerAsnIleArgAspLeuLeuProValLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysAlaGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnIleArgGlyIleLeuLysThrValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleAlaGluGluThrGlyLeuThrLeuGluLysAlaThrLeuAlaGluLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysArgIleGluValGlyLysGluAsnThrThrIleIleAspGlyAlaGlyGluAlaVal 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AsnIleGluAlaArgValLysGlnIleArgThrGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaArgThrAlaIleAlaAlaLeuThrGlyValAsnAlaAspGlnAsnAlaGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 LysIleValLeuArgAlaMetGluGluProLeuArgGlnIleValThrAsnGlyGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluAlaSerValValValAlaAlaValAlaAlaGlyLysGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaThrGlyGluTyrValAspMetValGluAlaGlyValValAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrAspAlaAla 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        521 ValAlaGluLeu------ProLysGluAspAlaProMetProGlyGlyMetProGlyGly 538

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
Db        539 MetGlyGlyMet---GlyMet 544

RESULT 8
CH60_BARHE
ID   CH60_BARHE     STANDARD;      PRT;   547 AA.
AC   O33963; O87267;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE   60).
GN   GROL OR GROEL OR MOPA.
OS   Bartonella henselae (Rochalimaea henselae).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bartonellaceae; Bartonella.
OX   NCBI_TaxID=38323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RX   MEDLINE=97419519; PubMed=9274034;
RA   Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;
RT   "Heat shock response and groEL sequence of Bartonella henselae and
RT   Bartonella quintana.";
RL   Microbiology 143:2807-2815(1997).
RN   [2]
RP   SEQUENCE OF 1-543 FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RA   Marston E.L., Sumner J.W., Regnery R.L.;
RT   "Evaluation of intraspecies genetic variation within the 60 kDa heat
RT   shock protein (groEL) gene of Bartonella species: a new phylogenetic
```

```
RT   analysis tool.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-408 FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RX   MEDLINE=97373904; PubMed=9230387;
RA   Sumner J.W., Nicholson W.L., Massung R.F.;
RT   "PCR amplification and comparison of nucleotide sequences from the
RT   groESL heat shock operon of Ehrlichia species.";
RL   J. Clin. Microbiol. 35:2087-2092(1997).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78514; AAB69094.1; -.
DR   EMBL; AF014829; AAD04238.1; ALT_INIT.
DR   EMBL; U96734; AAB65637.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   547 AA;  57625 MW;  EDE25D566D5B10D2 CRC64;
```

Alignment Scores:

Pred. No.:	4.89e-96	Length:	547
Score:	1867.50	Matches:	385
Percent Similarity:	81.24%	Conservative:	61
Best Local Similarity:	70.13%	Mismatches:	96
Query Match:	62.92%	Indels:	7
DB:	1	Gaps:	4

US-09-077-574A-1 (1-1647) x CH60_BARHE (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||||:::   |||      :::|||||||||:::|||   ||||||||
Db          1 MetAlaAlaLysGluValLysPheGlyArgGluAlaArgGluArgLeuLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||:::|||   ||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||:::::|||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||||||||||||||||||||||||||||||||||   |||||||||
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuGlyGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              :::|||||||||||   ||||||||||   |||||||||   :::||||||||||||
Db        101 GlnGluGlyValLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||   ||||||      |||         |||             |||      :::
Db        121 AspAlaAlaValAspGluValValAlaAsnLeuPheLysLysAlaLysLysIleGlnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                  ||||||||||||||||||||||||||||||||                ||||||
Db        141 SerAlaGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAlaAlaGluIleGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::||||||:::||||||   |||||||||       ||||||||||||||||||||||
Db        161 MetIleAlaAspAlaMetGluLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      ||||||   |||:::|||||||||||||||:::|||||||||||||||||||
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||   ||||||||||||   :::||||||:::|||||||||
Db        201 SerProTyrPheValThrAsnAlaGluLysMetValAlaAspLeuAspAspProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||||:::::::::::::::   :::||||||:::||||||   |||
Db        221 LeuIleHisGluLysLysLeuSerAsnLeuGlnSerLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::      :::||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnSerGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||   |||::::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||   |||:::
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrSerGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   |||||||||||||||||||:::|||      ||||||   |||
Db        301 IleSerGluAspValGlyIleLysLeuGluAsnValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   |||   |||||||||||||||||:::|||||||||||:::
Db        321 LysLysValAsnIleSerLysGluAsnThrThrIleIleAspGlyAlaGlyGlnLysSer 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              :::|||   |||||||||   ||||||:::   |||||||||||||||:::|||||||||
Db        341 GluIleAsnAlaArgValAsnGlnIleLysValGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   |||||||||   |||:::|||||||||||||||||||:::|||||||||
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAlaAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||   :::
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::         |||      :::|||   ::::::   |||:::   ||||||:::
Db        421 ArgAlaAlaAsnAlaLeu---ThrValLysGlySerAsnProAspGlnGluAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||:::||||||:::|||:::   |||   ||||||||||   |||||||||
Db        440 AsnIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaThrAsnAlaGlyGlu 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAA---GATGGTTTTGGATTTAAT 1437
              |||   :::|||:::|||   ||||||   |||         |||   ||||||:::|||
Db        460 GluAlaAlaIleIleValGlyLysValLeuGluAsnAsnAlaAspThrPheGlyTyrAsn 479
```

```
Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        480 ThrAlaThrGlyGluPheGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        500 ValArgSerAlaLeuGlnAsnAlaAlaSerIleAlaSerLeuLeuIleThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATG---------CCTATGCCTGGCGGT 1608
Db        520 MetValAlaGlu------ValProLysLysAspThrProValProProMetProGlyGly 537

Qy       1609 GGTATGGGTGGTATGGGTGGTATGGAC 1635
Db        538 GlyMetGlyGlyMetGlyGlyMetAsp 546
```

RESULT 9
CH61_RHIME
ID CH61_RHIME STANDARD; PRT; 545 AA.
AC P35469;
DT 01-JUN-1994 (Rel. 29, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1).
GN GROL1 OR GROEL1 OR GROEL-A OR R00792 OR SMC00913.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=93231539; PubMed=8097179;
RA Rusanganwa E., Gupta R.S.;
RT "Cloning and characterization of multiple groEL chaperonin-encoding
RT genes in Rhizobium meliloti.";
RL Gene 126:67-75(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA Ogawa J.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC

DR EMBL; M94192; AAA26285.1; -.
DR EMBL; U19726; AAA61955.1; -.
DR EMBL; AL591785; CAC45364.1; -.
DR PIR; JN0509; JN0509.
DR HSSP; P06139; 1GRL.
DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Heat shock; Multigene family;
KW Complete proteome.
FT CONFLICT 12 12 A -> G (IN REF. 1).
SQ SEQUENCE 545 AA; 57687 MW; 4CEF358957E1B45A CRC64;

Alignment Scores:

Pred. No.:	5.21e-96	Length:	545
Score:	1867.00	Matches:	388
Percent Similarity:	81.24%	Conservative:	58
Best Local Similarity:	70.67%	Mismatches:	97
Query Match:	62.90%	Indels:	6
DB:	1	Gaps:	4

US-09-077-574A-1 (1-1647) x CH61_RHIME (1-545)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaValAlaGluValValLysAspLeuLeuAlaLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerAspGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluLysGlnIleGlyLeu 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 AspIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240
```

```
Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnThrGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGC---TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db        421 ArgSerSerValLysIle------ThrValLysGlyGluAsnAspAspGlnAspAlaGly 438

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db        439 ValAsnIleValArgArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAA---GATGGTTTTGGATTT 1434
Db        459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspAspPheGlyTyr 478

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db        479 AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db        499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATG 1614
Db        519 AlaMetIleAlaGluLeu------ProLysLysAspAlaProAlaMetProGlyGlyMet 536

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db        537 GlyGlyMetGlyGlyMetAspMetMet 545

RESULT 10
CH60_AMOPS
ID   CH60_AMOPS     STANDARD;      PRT;   551 AA.
AC   P26004;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Amoeba proteus symbiotic bacterium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Legionellaceae.
OX   NCBI_TaxID=2728;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ahn T.I., Leeu H.K., Kwak I.H., Jeon K.W.;
RT   "Nucleotide sequence and temperature-dependent expression of groEL
RT   gene isolated from symbiotic bacteria of Amoeba proteus.";
RL   Endocyt. Cell Res. 8:33-44(1991).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M86549; AAC09381.1; -.
DR   PIR; JC2562; JC2562.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   551 AA;  58332 MW;  A25895CBF9585DA0 CRC64;

Alignment Scores:
Pred. No.:              5.92e-96        Length:         551
Score:                  1866.00         Matches:        373
Percent Similarity:     81.42%          Conservative:   74
Best Local Similarity:  67.94%          Mismatches:     96
Query Match:            62.87%          Indels:         6
DB:                     1               Gaps:           2

US-09-077-574A-1 (1-1647) x CH60_AMOPS (1-551)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db          2 AlaLysGluLeuArgPheGlyAspAspAlaArgGlnGlnMetLeuAlaGlyValAsnAla 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db         22 LeuAlaAspArgValLysAlaThrMetGlyProSerGlyArgAsnValValLeuGluArg 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db         42 SerPheGlyAlaProThrValThrLysAspGlyValSerValAlaLysGluIleGluPhe 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db         62 GluAsnArgPheLysAsnMetGlyAlaGlnMetValLysGluValAlaAlaLysThrSer 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db         82 AspThrAlaGlyAspGlyThrThrThrAlaThrValLeuAlaArgSerIleValValGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        102 GlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIleAspLys 121
```

```
Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        122 AlaValThrAlaIleThrLysGluLeuGlnLysMetSerLysProCysLysAspGlyLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        142 AlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGlnAlaIleGlySerIleIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        162 AlaGluAlaMetGluLysValGlyLysGluGlyValIleThrValGluAspGlyAsnGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db        182 LeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        202 TyrPheIleAsnAsnGlnGlnAsnMetSerAlaGluLeuGluHisProPheIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        222 ValAspLysLysIleAlaThrIleArgAspMetLeuSerValLeuGluAlaValAlaLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        242 SerGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db        262 ValValAsnAsnMetArgGlyIleValLysValCysAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db        282 AspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrAsnGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db        302 GluGluIleGlyThrSerLeuGluThrAlaSerLeuGluSerLeuGlyThrAlaLysArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db        322 IleValValThrLysGluAsnThrThrIleIleAspGlyGluGlyLysAlaThrGluIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db        342 AsnAlaArgIleAlaGlnIleArgAlaGlnMetGluGluThrSerSerAspTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db        362 GluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db        382 AlaAlaThrGluIleGluMetLysGluLysLysAlaArgValGluAspAlaLeuHisAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db        402 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyValAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db        422 GlnLysValLeuAspGlyLeuLysGlyAspAsnAlaAspGlnAspMetGlyIleAsnIle 441

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db        442 LeuArgArgAlaIleGluSerProLeuArgGlnIleValAlaAsnAlaGlyTyrGluSer 461

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db        462 SerValIleValAsnLysValAlaGluHisLysAspAsnPheGlyPheAsnAlaAlaThr 481

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db        482 GlyGlnTyrGlyAspMetValGluMetGlyIleLeuAspProThrLysValThrArgThr 501

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db        502 AlaLeuGlnAsnAlaAlaSerValArgSerLeuMetLeuThrThrGluCysMetValAla 521

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC----------GGTGGTATG 1614
Db        522 AspLeu------ProLysLysAspGluGlyMetAlaGlyAlaGlyAspMetGlyGlyMet 539

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db        540 GlyGlyMetGlyGlyMetGlyGlyMet 548
```

RESULT 11
CH60_BARQU
ID   CH60_BARQU   STANDARD;   PRT;   547 AA.
AC   O33964;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE   60).
GN   GROL OR GROEL OR MOPA.
OS   Bartonella quintana (Rochalimaea quintana).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bartonellaceae; Bartonella.
OX   NCBI_TaxID=803;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51694;
RX   MEDLINE=97419519; PubMed=9274034;
RA   Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;
RT   "Heat shock response and groEL sequence of Bartonella henselae and
RT   Bartonella quintana.";
RL   Microbiology 143:2807-2815(1997).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC   proper assembly of unfolded polypeptides generated under stress
CC   conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC   7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U78515; AAB69095.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   547 AA; 57610 MW; 8620CFC50BF2D6F4 CRC64;

Alignment Scores:

Pred. No.:	7.15e-96	Length:	547
Score:	1864.50	Matches:	385
Percent Similarity:	81.24%	Conservative:	61
Best Local Similarity:	70.13%	Mismatches:	96
Query Match:	62.82%	Indels:	7
DB:	1	Gaps:	4

US-09-077-574A-1 (1-1647) x CH60_BARQU (1-547)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyArgGluAlaArgGluArgLeuLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AspIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuGlyGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 GlnGluGlyValLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspAlaAlaValGluGluValValGlyAsnLeuPheLysLysAlaLysLysIleGlnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerAlaGluIleAlaGlnValGlyThrIleSerAlaAsnGlyAlaAlaGluIleGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 MetIleAlaAspAlaMetGluLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuAspAspProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuIleHisGluLysLysLeuSerAsnLeuGlnSerLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnSerGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrSerGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspValGlyIleLysLeuGluAsnValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValAsnIleSerLysGluAsnThrThrIleIleAspGlyAlaGlyLysLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleAsnAlaArgValAsnGlnIleLysValGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaAlaAsnAlaLeu---AlaIleLysGlySerAsnProAspGlnGluAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        440 AsnIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaThrAsnAlaGlyGlu 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAA---GATGGTTTTGGATTTAAT 1437
Db        460 GluAlaAlaIleIleValGlyLysValLeuGluAsnAsnAlaAspThrPheGlyTyrAsn 479

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        480 ThrAlaThrGlyGlnPheGlyAspLeuIleAlaLeuGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        500 ValArgSerAlaLeuGlnAsnAlaAlaSerIleAlaSerLeuLeuIleThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATG---------CCTATGCCTGGCGGT 1608
Db        520 MetValAlaGlu------ValProLysLysAspThrProMetProProMetProGlyGly 537

Qy       1609 GGTATGGGTGGTATGGGTGGTATGGAC 1635
Db        538 GlyMetGlyGlyMetGlyGlyMetAsp 546
```

RESULT 12
CH64_RHIME
ID CH64_RHIME STANDARD; PRT; 545 AA.
AC Q92ZQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin 4 (Protein Cpn60 4) (groEL protein 4).
GN GROL4 OR GROEL4 OR RA0395 OR SMA0744.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

```
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE007230; AAK65053.1; -.
DR   PIR; C95311; C95311.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
KW   Complete proteome.
SQ   SEQUENCE   545 AA;  57701 MW;  49BA61F23CE70431 CRC64;
```

Alignment Scores:

Pred. No.:	7.62e-96	Length:	545
Score:	1864.00	Matches:	387
Percent Similarity:	81.24%	Conservative:	59
Best Local Similarity:	70.49%	Mismatches:	97
Query Match:	62.80%	Indels:	6
DB:	1	Gaps:	4

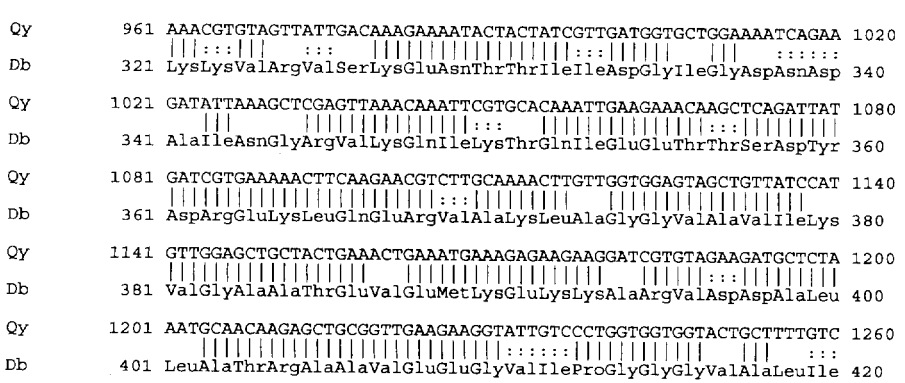
US-09-077-574A-1 (1-1647) x CH64_RHIME (1-545)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyArgSerAlaArgGluLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ArgGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaValAlaGluValValLysAspLeuLeuAlaLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerAspGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluLysGlnIleGlyLeu 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 AspIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnThrGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleThrLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysSer 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AspIleGluGlyArgValAlaGlnIleLysAlaGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgIleAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGlnGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGC---TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db        421 ArgSerSerValLysIle------ThrValLysGlyGluAsnAspAspGlnAspAlaGly 438

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db        439 ValAsnIleValArgArgAlaLeuGlnSerProAlaArgGlnIleValGluAsnAlaGly 458

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAA---GATGGTTTTGGATTT 1434
Db        459 AspGluAlaSerIleValValGlyLysIleLeuGluLysAsnThrAspAspPheGlyTyr 478

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
Db        479 AsnAlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleIleAspProValLys 498

Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db        499 ValValArgThrAlaLeuGlnAspAlaAlaSerValAlaSerLeuLeuIleThrThrGlu 518

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATG 1614
Db        519 AlaMetIleAlaGluLeu------ProLysLysAspAlaProAlaMetProGlyGlyMet 536

Qy       1615 GGTGGTATGGGTGGTATGGACGGTATG 1641
Db        537 GlyGlyMetGlyGlyMetAspMetMet 545
```

RESULT 13

```
CH60_XYLFA
ID   CH60_XYLFA     STANDARD;      PRT;   547 AA.
AC   Q9PFP2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR XF0615.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE003907; AAF83425.1; -.
DR   PIR; F82783; F82783.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Complete proteome.
SQ   SEQUENCE   547 AA;  57757 MW;  30BE7F937CA7A9D8 CRC64;
```

```
Alignment Scores:
Pred. No.:              9.22e-96        Length:         547
Score:                  1862.50         Matches:        366
Percent Similarity:     81.39%          Conservative:   80
Best Local Similarity:  66.79%          Mismatches:     97
Query Match:            62.75%          Indels:         5
DB:                     1               Gaps:           2

US-09-077-574A-1 (1-1647) x CH60_XYLFA (1-547)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              |||||:::|||||||||:::|||      |||||||||   ::::::      ||||||
Db          1 MetAlaAlaLysGluIleIlePheSerGluLysAlaArgSerArgMetValHisGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              :::   ||||||||||||||||||   ||||||||||||||||||||:::||||||:::
Db         21 AsnLeuLeuAlaAsnAlaValLysAlaThrLeuGlyProLysGlyArgHisValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlySerProIleIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||   ||||||||||||||||||||||||||||:::|||||||||||||   |||
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||   ||||||||||||||||||||||||||||||||||||||||||:::
Db         81 ThrAsnAspHisAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaLeuIle 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 ArgGluGlyCysLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||:::|||          ||||||      ||:::|||||||||   |||
Db        121 AspLysAlaValIleAlaAlaValThrGluLeuLysLysIleSerLysProThrSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||   ||||||||||||   ||||||||||||||||||||   :::|||||||||||
Db        141 AspLysAlaIleAlaGlnValAlaThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||   ||||||||||||   ||||||||||||:::||||||
Db        161 IleIleAlaGluAlaMetLysLysValGlyLysGluGlyValIleThrIleGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||      ||||||||||||||||||||||:::||||||||||||||||
Db        181 ThrThrLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrSer 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||:::   |||   :::      :::   ||||||||||||||||||||
Db        201 SerProTyrPheIleAsnAsnGlnGlnSerGlnIleValGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   ::::::||||||||:::|||:::::::|||:::|||   :::|||:::   |||
Db        221 LeuLeuHisAspLysLysIleSerSerValArgAspLeuLeuThrValLeuAspAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||   ::::::||||||||||:::|||||:::||||||||||||||||||||||
Db        241 AlaLysGluSerLysProLeuLeuIleValAlaGluGluValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||   :::||||||   ::::::|||   ||||||||||||||||||
Db        261 ThrLeuValValAsnAsnIleArgGlyIleIleLysValCysAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||:::|||:::||||||||||||
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspMetAlaValLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   |||:::   |||:::   ||||||      :::   |||   ||||||   |||
Db        301 IleSerGluGluValGlyLeuSerLeuGluLysAlaThrThrSerHisLeuGlyLysAla 320
```



```
Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValArgValSerLysGluAsnThrThrIleIleAspGlyIleGlyAspAsnAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 AlaIleAsnGlyArgValLysGlnIleLysThrGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 LeuAlaThrArgAlaAlaValGluGluGlyValIleProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaIleThrAlaIleSerAsnLeuLysGlyAlaAsnGluAspGlnThrHisGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 GlnIleAlaLeuArgAlaMetGluAlaProLeuArgGluIleValAlaAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluProSerValIleLeuAsnLysValLysGluGlyLysAspAsnPheGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 AlaThrGlyGluPheGlyAspMetValAsnLeuGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnAsnAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCT---------GGCGGTGGT 1611
Db        521 ValAlaGlu------AlaProLysLysAspGluProThrProProAlaAlaGlyGlyGly 538

Qy       1612 ATGGGTGGTATGGGTGGTATGGAC 1635
Db        539 MetGlyGlyMetGlyGlyMetAsp 546
```

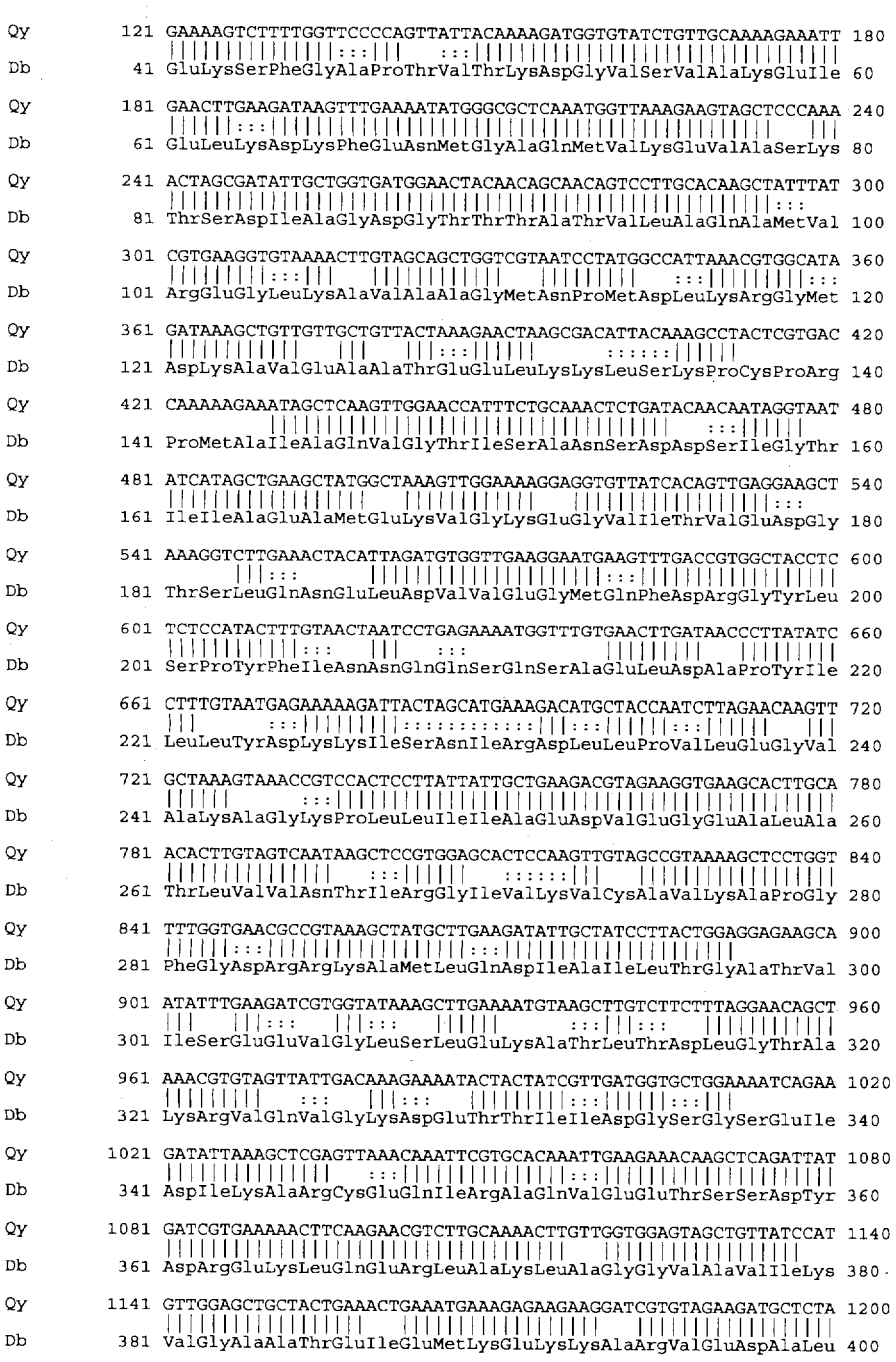
```
RESULT 14
CH60_RHOMR
ID   CH60_RHOMR     STANDARD;      PRT;   540 AA.
AC   Q9XCA9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Rhodothermus marinus (Rhodothermus obamensis).
OC   Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC   Crenotrichaceae; Rhodothermus.
OX   NCBI_TaxID=29549;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ITI 376;
RA   Thorolfsdottir E.T.T., Backman V.M., Blondal T.,
RA   Thorbjarnardottir S.H., Palsdottir A., Hauksdottir H.,
RA   Kristjansdottir S., Eggertsson G.;
RT   "Heat shock in Rhodothermus marinus: cloning and sequence analysis of
RT   the groESL, dnaK and dnaJ genes.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF145252; AAD37976.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   540 AA;  57666 MW;  6867448B7BEF18EC CRC64;
```

Alignment Scores:			
Pred. No.:	3.28e-95	Length:	540
Score:	1852.50	Matches:	371
Percent Similarity:	82.46%	Conservative:	71
Best Local Similarity:	69.22%	Mismatches:	93
Query Match:	62.42%	Indels:	1
DB:	1	Gaps:	1

US-09-077-574A-1 (1-1647) x CH60_RHOMR (1-540)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGlnIleThrPheAsnAlaAspAlaArgMetAlaLeuLysArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AspLysLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValIleIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 GluLysLysPheGlyAlaProThrValThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysLeuGluAsnValGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrSerAspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleLeu 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ThrAlaGlyLeuLysSerValThrAlaGlyAlaAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValGluValValValAlaGluLeuArgLysMetSerGlnGluValGlnAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 LysAsnArgIleAlaGlnValAlaThrIleSerAlaAsnGlyAspLysAlaIleGlyGln 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 LeuIleAlaAspAlaPheGluLysValGlyLysAspGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
```

```
Db        181 LysGlyThrGluThrThrLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db        201 SerProTyrPheValThrAsnProAspThrMetGluAlaValLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db        221 LeuIleHisAspLysLysIleSerAlaMetLysAspLeuLeuProIleLeuGluLysVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db        241 ValGlnThrGlyArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db        261 ThrLeuValValAsnLysLeuArgGlyValLeuLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db        301 IleSerGluGluLysGlyTyrArgLeuGluAsnAlaThrLeuAspTyrLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db        321 GluArgIleIleValAspLysAspAsnThrThrIleValGlyLysGlyAspProAla 340

Qy       1021 GATATTAAAGCTCCAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db        341 GlnIleLysAlaArgAlaAsnGlnIleArgGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValLeuLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db        381 IleGlyAlaAlaThrGluProGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaTyrIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db        421 ArgAlaIleAlaAlaLeuAspLysValGluValGluAsnGluAspGlnLysIleGlyVal 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db        441 GlnIleValGlnArgAlaLeuGluProLeuArgGlnIleAlaAlaAsnAlaGlyTrp 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db        461 GluGlySerIleValValGlnArgValLysGluGlyGlnGlyAspPheGlyPheAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db        481 GlnThrGluGluPheGlyAsnLeuLeuGluGlnGlyValIleAspProThrLysValAla 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db        501 ArgThrAlaLeuGluAsnAlaAlaSerValAlaGlyLeuLeuLeuThrThrGluAlaVal 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAA---GATATGCCTATGCCTGGC 1605

Db        521 ValAlaGluLysProGluLysGluLysAlaAlaAlaProSerProGly 536

RESULT 15
CH60_CHRVI
ID   CH60_CHRVI     STANDARD;      PRT;   546 AA.
AC   P31293;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Chromatium vinosum.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC   Chromatiaceae; Allochromatium.
OX   NCBI_TaxID=1049;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93186721; PubMed=8444812;
RA   Ferreyra R., Soncini F., Viale A.M.;
RT   "Cloning, characterization, and functional expression in Escherichia
RT   coli of chaperonin (groESL) genes from the phototrophic sulfur
RT   bacterium Chromatium vinosum.";
RL   J. Bacteriol. 175:1514-1523(1993).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=99008942; PubMed=9790891;
RA   Dionisi H.M., Viale A.M.;
RT   "Purification and characterization of Chromatium vinosum GroEL and
RT   GroES proteins overexpressed in Escherichia coli cells lacking the
RT   endogenous groESL operon.";
RL   Protein Expr. Purif. 14:275-282(1998).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions.
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M99443; AAA23319.1; -.
DR   PIR; B47073; B47073.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57541 MW;  3250141881C04DD6 CRC64;

Alignment Scores:
Pred. No.:              3.28e-95        Length:         546
Score:                  1852.50         Matches:        374
Percent Similarity:     81.02%          Conservative:   70
Best Local Similarity:  68.25%          Mismatches:     101
Query Match:            62.42%          Indels:         3
DB:                     1               Gaps:           2

US-09-077-574A-1 (1-1647) x CH60_CHRVI (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db          1 MetSerAlaLysAspValLysPheGlyGlyAspAlaArgValArgMetMetGluGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120

Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40
```



Search completed: January 28, 2004, 13:26:11
Job time : 45.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 28, 2004, 13:21:36 ; Search time 72 Seconds
(without alignments)
11805.907 Million cell updates/sec

Title: US-09-077-574A-1
Perfect score: 2968
Sequence: 1 atggcttctaaagaaatcct.....gtatggacggtatgtactag 1647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09077574/runat_28012004_125846_11617/app_query.fasta_1.1799
-DB=SPTREMBL_23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09077574_@CGN_1_1_92_@runat_28012004_125846_11617 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2708	91.2	548	2	O87888	O87888 lawsonia in
2	1880	63.3	546	2	Q8GBD2	Q8gbd2 acetobacter
3	1823	61.4	552	2	Q8KJ46	Q8kj46 rhizobium l
4	1819	61.3	546	16	Q8FX87	Q8fx87 brucella su
5	1804.5	60.8	545	2	Q8GB95	Q8gb95 heliobacill
6	1796.5	60.5	549	2	Q9AJB5	Q9ajb5 pseudoalter
7	1780.5	60.0	545	16	Q8CX48	Q8cx48 shewanella
8	1780.5	60.0	546	2	Q8GBB4	Q8gbb4 azotobacter
9	1768	59.6	539	2	Q9RC20	Q9rc20 bacillus sp
10	1768	59.6	546	2	O31198	O31198 leptospira
11	1754	59.1	539	2	Q9EZV4	Q9ezv4 bacillus st
12	1742	58.7	545	2	Q93FU8	Q93fu8 enterobacte
13	1736	58.5	538	2	O33688	O33688 primary end
14	1722.5	58.0	546	2	Q8RNU2	Q8rnu2 helicobacte
15	1714.5	57.8	548	2	Q8KIW5	Q8kiw5 buchnera sp
16	1711	57.6	543	2	Q8KZN2	Q8kzn2 thermus sp.
17	1710.5	57.6	548	2	Q8KIW7	Q8kiw7 buchnera sp
18	1706	57.5	555	2	Q9X603	Q9x603 primary end
19	1694	57.1	542	16	Q8CWW6	Q8cww6 streptococc
20	1691.5	57.0	536	2	Q8KIW2	Q8kiw2 buchnera sp
21	1691	57.0	537	2	Q8KJ14	Q8kj14 streptococc
22	1687	56.8	545	16	Q8CXL3	Q8cxl3 oceanobacil
23	1686	56.8	540	16	Q8CX22	Q8cx22 streptococc
24	1685	56.8	531	2	Q8GBD0	Q8gbd0 enterococcu
25	1685	56.8	540	16	Q8CX00	Q8cx00 streptococc
26	1683.5	56.7	537	2	Q8KIW4	Q8kiw4 buchnera sp
27	1681	56.6	536	2	Q8KJ10	Q8kj10 streptococc
28	1679.5	56.6	537	2	Q8KIW3	Q8kiw3 buchnera sp
29	1676.5	56.5	537	2	Q8KIW9	Q8kiw9 buchnera sp
30	1675	56.4	534	2	Q8GBB6	Q8gbb6 streptococc
31	1672.5	56.4	548	2	O08499	O08499 rhopalosiph
32	1671.5	56.3	548	2	O08500	O08500 sitobion av
33	1670	56.3	531	2	Q8GBC6	Q8gbc6 enterococcu
34	1669	56.2	531	2	Q8GBC4	Q8gbc4 enterococcu
35	1669	56.2	531	2	Q8GBC0	Q8gbc0 enterococcu
36	1668	56.2	534	2	Q8KIY3	Q8kiy3 streptococc
37	1667	56.2	534	2	Q8KJ12	Q8kj12 streptococc
38	1663.5	56.0	540	2	Q9X4R5	Q9x4r5 streptococc
39	1663	56.0	531	2	Q8GBC2	Q8gbc2 enterococcu
40	1663	56.0	535	2	Q8KJ08	Q8kj08 streptococc
41	1660	55.9	503	2	O87271	O87271 bartonella
42	1660	55.9	532	2	Q8GBC8	Q8gbc8 enterococcu
43	1659	55.9	532	2	Q8GBB8	Q8gbb8 enterococcu
44	1648.5	55.5	539	2	Q8GJ00	Q8gj00 fusobacteri
45	1644	55.4	532	16	Q8CWJ0	Q8cwj0 vibrio vuln

ALIGNMENTS

RESULT 1
O87888
ID O87888 PRELIMINARY; PRT; 548 AA.
AC O87888;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE GroEL/HSP60 homolog (60 kDa chaperonin) (Protein Cpn60) (groEL
DE protein).
OS Lawsonia intracellularis.
OG Plasmid pISI-2.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Lawsonia.
OX NCBI_TaxID=29546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98386497; PubMed=9720028;
RA Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D.,
RA Strugnell R.A.;
RT "Identification and sequencing of the groE operon and flanking genes
RT of Lawsonia intracellularis: use in phylogeny.";
RL Microbiology 144:0-0(0).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).

```
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; U45241; AAC36500.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone; Plasmid.
SQ   SEQUENCE   548 AA;  58605 MW;  6388C431E663E498 CRC64;

Alignment Scores:
Pred. No.:              4.44e-134       Length:         548
Score:                  2708.00         Matches:        548
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            91.24%          Indels:         0
DB:                     2               Gaps:           0

US-09-077-574A-1 (1-1647) x O87888 (1-548)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaSerLysGluIleLeuPheAspAlaLysAlaArgGluLysLeuSerArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspLysLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 GluLysSerPheGlySerProValIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaProLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrSerAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleTyr 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgGluGlyValLysLeuValAlaAlaGlyArgAsnProMetAlaIleLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspLysAlaValValAlaValThrLysGluLeuSerAspIleThrLysProThrArgAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlnLysGluIleAlaGlnValGlyThrIleSerAlaAsnSerAspThrThrIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleIleAlaGluAlaMetAlaLysValGlyLysGlyGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysGlyLeuGluThrThrLeuAspValValGluGlyMetLysPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetValCysGluLeuAspAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 LeuCysAsnGluLysLysIleThrSerMetLysAspMetLeuProIleLeuGluGlnVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AlaLysValAsnArgProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyAlaLeuGlnValValAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheGlyGluArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluAla 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IlePheGluAspArgGlyIleLysLeuGluAsnValSerLeuSerSerLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LysArgValValIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysSerGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspIleLysAlaArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuValGlyGlyValAlaValIleHis 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ValGlyAlaAlaThrGluThrGluMetLysGluLysLysAspArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaPheVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ArgSerIleLysValLeuAspAspIleLysProAlaAspAspAspGluLeuAlaGlyLeu 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AsnIleIleArgArgSerLeuGluGluProLeuArgGlnIleAlaAlaAsnAlaGlyTyr 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GluGlySerIleValValGluLysValArgGluProLysAspGlyPheGlyPheAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AlaSerGlyGluTyrGluAspLeuIleLysAlaGlyValIleAspProLysLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 ArgIleAlaLeuGlnAsnAlaAlaSerValAlaSerLeuLeuLeuThrThrGluCysAla 520

Qy       1561 ATTGCTGAAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 IleAlaGluLysProGluProLysLysAspMetProMetProGlyGlyGlyMetGlyGly 540

Qy       1621 ATGGGTGGTATGGACGGTATGTAC 1644
              ||||||||||||||||||||||||
Db        541 MetGlyGlyMetAspGlyMetTyr 548

RESULT 2
Q8GBD2
ID   Q8GBD2      PRELIMINARY;      PRT;   546 AA.
AC   Q8GBD2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL protein.
GN   GROEL.
OS   Acetobacter aceti.
```

```
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC   Acetobacteraceae; Acetobacter.
OX   NCBI_TaxID=435;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO3283;
RA   Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
RA   Yanagida F.;
RT   "Cloning and Characterization of groEL Operon in Acetobacter aceti.";
RL   J. Biosci. Bioeng. 94:140-147(2002).
DR   EMBL; AB081586; BAC16232.1; -.
SQ   SEQUENCE   546 AA;  58106 MW;  3E476AF42F2CD5CC CRC64;

Alignment Scores:
Pred. No.:              9.17e-91      Length:          546
Score:                  1880.00       Matches:         381
Percent Similarity:     82.42%        Conservative:    69
Best Local Similarity:  69.78%        Mismatches:      94
Query Match:            63.34%        Indels:          2
DB:                     2             Gaps:            2

US-09-077-574A-1 (1-1647) x Q8GBD2 (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::   |||   |||   ||||||:::::::::       ||||||
Db          1 MetAlaAlaLysAspValLysPheGlyAlaAspAlaArgGlnArgMetArgAlaGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||   ||||||:::|||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||:::|||   ||||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||   |||||||||||||||||||||||||||||:::::|||||||||   |||
Db         61 GluLeuAlaAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              |||||||||   |||   ||||||||||||   ||||||||   :::||||||||||||
Db        101 ArgGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              ||||||||||||      |||   :::||||||            |||
Db        121 AspLysAlaValAlaValValIleGluGluLeuLysLysAsnAlaLysLysValThrThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              |||   |||||||||||||||||||||||||||||   ::::::   ||||||
Db        141 ProAlaGluThrAlaGlnValGlyThrIleSerAlaAsnGlyGluSerGluIleGlyGln 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::|||:::|||||||||   |||||||||      ||||||||||||||||||||||||
Db        161 MetIleSerGluAlaMetGlnLysValGlySerGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      :::|||   |||||||||||||||||||:::|||||||||||||||:::
Db        181 LysHisPheGlnThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrIle 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||      :::|||:::||||||||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetThrAlaAspLeuGluAsnProTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||||:::::::|||::::::   |||||||||:::||||||   |||
Db        221 LeuIleHisGluLysLysLeuSerSerLeuGlnProMetLeuProLeuLeuGluSerVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
                 :::      ||||||||||||||||||||||||||||||:::||||||||||||||||
Db        241 ValGlnSerGlyArgProLeuLeuIleIleAlaGluAspValAspGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||   |||::::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||      ||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgAsnValMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   ||||||||||||||   |||:::||:::   ||||||||||||
Db        301 IleSerGluAspLeuGlyIleLysLeuGluThrValThrLeuAsnMetLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   |||||||||||||||||||||||||||||||||||||||||||:::::
Db        321 LysLysValHisIleAspLysGluAsnThrThrIleValAspGlyAlaGlyLysAlaAsp 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              |||||||||   |||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspIleLysGlyArgValLysGlnIleArgAlaGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   :::|||||   |||:::||||||:::|||||||||||:::|||||||||
Db        381 ValGlyGlySerThrGluValGluValLysGluArgLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 HisAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaLeuAla 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
              |||:::        |||:::  :::        :::||||||:::      |||
Db        421 ArgAlaThrLeuLysLeuGluGlyLeuHisTyrHisAsnAspAspGlnArgValGlyGly 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
              :::|||||||||||:::|||:::   |||||||||||||||||   |||||||||
Db        441 AspIleIleArgArgAlaLeuGlnAlaProLeuArgGlnIleAlaHisAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGAT---GGTTTTGGATTTAAT 1437
              :::|||:::::::::      ||||||   |||      |||      ||||||||:::
Db        461 AspGlyAlaValIleAlaAsnLysValLeuGluAsnSerAspTyrAsnPheGlyPheAsp 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
              |||   :::|||||||||:::::|||::::::||||||:::||||||||   ||||||
Db        481 AlaGlnAlaGlyGluTyrLysAsnLeuValGluAlaGlyIleIleAspProAlaLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
                 |||   |||||||||:::|||||||||||||||   ||||||:::|||||||||
Db        501 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuIleThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
                 :::||||||:::||||||   ||||||      |||      |||      ||||||
Db        521 MetValAlaGluArgProGlu---LysLysAlaProAlaGlyGlyProAspMetGly 539

Qy       1618 GGTATGGGTGGTATGGAC 1635
              ||||||||||||||||||
Db        540 GlyMetGlyGlyMetAsp 545

RESULT 3
Q8KJ46
ID   Q8KJ46      PRELIMINARY;      PRT;   552 AA.
AC   Q8KJ46;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable chaperonin groEL DF protein (60 kDa chaperonin) (Protein
DE   Cpn60) (groEL protein).
GN   GROEL.
OS   Rhizobium loti (Mesorhizobium loti).
```

```
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R7A;
RX   MEDLINE=21999272; PubMed=12003951;
RA   Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,
RA   Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,
RA   Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;
RT   "Comparative sequence analysis of the symbiosis island of
RT   Mesorhizobium loti strain R7A.";
RL   J. Bacteriol. 184:3086-3095(2002).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AL672115; CAD31231.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   552 AA;  57747 MW;  7D0A8C53CD7FA6AB CRC64;

Alignment Scores:
Pred. No.:                8.8e-88        Length:          552
Score:                    1823.00        Matches:         374
Percent Similarity:       80.98%         Conservative:    73
Best Local Similarity:    67.75%         Mismatches:      97
Query Match:              61.42%         Indels:          8
DB:                       2              Gaps:            4

US-09-077-574A-1 (1-1647) x Q8KJ46 (1-552)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheSerArgAspAlaArgGluArgMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValThrValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 ThrAsnAspIleAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 GlnGluGlyHisLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLeuAlaValThrAspValValAlaThrLeuIleLysAsnAlaLysLysIleLysThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerGluGluValAlaGlnValGlyThrIleAlaGlyAsnGlyAspGluSerValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db        161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheValThrAsnAlaAspLysMetValAlaAspLeuGluAspValTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnThrMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 ValGlnThrSerLysProLeuLeuIleIleSerGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluAspLeuGlyIleLysLeuGluAsnValGlyLeuAsnMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyLysLysGlu 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GluIleGlnGlyArgValAlaGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspLysGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyValAlaLeuLeu 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgAlaSerLeuSerIleAsnAlaVal---GlyAlaAsnSerAspGlnAlaAlaGlyIle 439

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        440 AsnIleValArgArgAlaLeuGlnAlaProAlaArgGlnIleAlaAlaAsnAlaGlyAla 459

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGT---TTTGGATTTAAT 1437
Db        460 GluAlaSerIleValAlaGlyLysIleLeuGluAsnLysGlyAlaThrPheGlyTyrAsn 479

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        480 AlaGlnThrGlyGluTyrGlyAspMetIleAlaMetGlyIleValAspProValLysVal 499

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        500 ValArgThrAlaLeuGlnAspAlaAlaSerValAlaGlyLeuLeuValThrThrGluAla 519

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAA----------GATATGCCT------ATG 1599
Db        520 MetIleAlaGluAlaProLysLysGluSerAlaGlyGlyGlyGlyMetProGlyGlyMet 539
```

```
Qy       1600 CCTGGCGGTGGTATGGGTGGTATGGGTGGTATGGAC 1635
              ||||||||||||||||||||||||||||||||||
Db        540 GlyGlyGlyGlyMetGlyGlyMetGlyGlyMetAsp 551

RESULT 4
Q8FX87
ID   Q8FX87      PRELIMINARY;      PRT;   546 AA.
AC   Q8FX87;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Chaperonin, 60 kDa.
GN   GROEL OR BRA0195.
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR   EMBL; AE014521; AAN33401.1; -.
DR   TIGR; BRA0195; -.
KW   Complete proteome.
SQ   SEQUENCE   546 AA;  57501 MW;  2E4046EDC121A30A CRC64;

Alignment Scores:
Pred. No.:              1.42e-87        Length:         546
Score:                  1819.00         Matches:        377
Percent Similarity:     80.33%          Conservative:   64
Best Local Similarity:  68.67%          Mismatches:     100
Query Match:            61.29%          Indels:         8
DB:                     16              Gaps:           5

US-09-077-574A-1 (1-1647) x Q8FX87 (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
              ||||||:::|||:::::::   |||          ||||||||||||:::   ||||||||
Db          1 MetAlaAlaLysAspValLysPheGlyArgThrAlaArgGluLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
              |||   ||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db         21 AspIleLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValIle 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
              :::|||||||||||:::|||   |||||||||||||||||||||||||||||||||:::
Db         41 AspLysSerPheGlyAlaProArgIleThrLysAspGlyValSerValAlaLysGluVal 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
              ||||||||||||||||||||||||||||||||||||||:::::|||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetLeuArgGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
              |||:::|||   |||||||||||||||||||||||||||||||   |||||||||
Db         81 ThrAsnAspThrAlaGlyAspGlyThrThrThrAlaThrValLeuGlyGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
              :::||||||   |||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 GlnGluGlyAlaLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
              |||   ||||||   |||   ||||||   ||||||||   |||
Db        121 AspLeuAlaValAsnGluValValAlaGluLeuLeuLysLysAlaLysLysIleAsnThr 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
              :::|||:::|||||||||||||||||||||||||||||   :::      ||||||
Db        141 SerGluGluValAlaGlnValGlyThrIleSerAlaAsnGlyGluAlaGluIleGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
              :::|||||||||||||||   |||||||||      ||||||||||||||||||||||||
Db        161 MetIleAlaGluAlaMetGlnLysValGlyAsnGluGlyValIleThrValGluGluAla 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
              |||      ||||||   |||:::|||||||||||||||:::|||||||||||||||||
Db        181 LysThrAlaGluThrGluLeuGluValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
              ||||||||||||||||||||||||||||||||   :::|||:::::   ||||||
Db        201 SerProTyrPheValThrAsnProGluLysMetValAlaAspLeuGluAspAlaTyrIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
              |||   :::|||||||||:::::::::::::::   :::||||||:::||||||   |||
Db        221 LeuLeuHisGluLysLysLeuSerAsnLeuGlnAlaLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
              :::   ::::::|||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ValGlnThrSerLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
              ||||||||||||||||||||||||||||   |||:::::   ||||||||||||||||||
Db        261 ThrLeuValValAsnLysLeuArgGlyGlyLeuLysIleAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||:::
Db        281 PheGlyAspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
              |||   ||||||   ||||||||||||||:::|||:::|||      ||||||   |||
Db        301 IleSerGluAspLeuGlyIleLysLeuGluSerValThrLeuAspMetLeuGlyArgAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
              |||:::|||   |||   ||||||||||||||||||||||||||||||||||:::
Db        321 LysLysValSerIleSerLysGluAsnThrThrIleValAspGlyAlaGlyGlnLysAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
              :::|||   ||||||||   ||||||:::   ||||||||||||||:::|||||||||
Db        341 GluIleAspAlaArgValGlyGlnIleLysGlnGlnIleGluGluThrThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleArg 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
              ||||||   ||||||||   |||:::|||||||||||||||||||:::|||||||||
Db        381 ValGlyGlyAlaThrGluValGluValLysGluLysLysAspArgValAspAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
              ||||||||||||||||||||||||||||||||   ||||||||||||||   :::
Db        401 AsnAlaThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuLeu 420

Qy       1261 CGC---TCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
              |||   |||   |||:::          |||      :::   |||:::   ||||||
Db        421 ArgAlaSerThrLysIle------ThrAlaLysGlyValAsnAlaAspGlnGluAlaGly 438

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
              :::||||||:::||||||:::::::::   |||   |||||||||      |||||||||
Db        439 IleAsnIleValArgArgAlaIleGlnAlaProAlaArgGlnIleThrThrAsnAlaGly 458

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAA---CCAAAAGATGGTTTTGGATTT 1434
              |||   |||:::::|||   |||:::   |||          :::   ||||||:::
Db        459 GluGluAlaSerValIleValGlyLysIleLeuGluAsnThrSerGluThrPheGlyTyr 478

Qy       1435 AATGCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAA 1494
              |||   |||:::|||||||||   |||||||||      |||::::::||||||   |||
Db        479 AsnThrAlaAsnGlyGluTyrGlyAspLeuIleSerLeuGlyIleValAspProValLys 498
```

```
Qy       1495 GTTACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAA 1554
Db        499 ValValArgThrAlaLeuGlnAsnAlaAlaSerValAlaGlyLeuLeuIleThrThrGlu 518

Qy       1555 TGCGCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGT 1608
Db        519 AlaMetIleAlaGluLeu------ProLysLysAspAlaAlaProAlaGlyMetProGly 536

Qy       1609 GGTATGGGTGGTATGGGTGGTATGGAC 1635
Db        537 GlyMetGlyGlyMetGlyGlyMetAsp 545

RESULT 5
Q8GB95
ID   Q8GB95      PRELIMINARY;      PRT;   545 AA.
AC   Q8GB95;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   60 kDa chaperonin groEL (Fragment).
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;
OC   Heliobacillus.
OX   NCBI_TaxID=28064;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22337798; PubMed=12446909;
RA   Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA   Blankenship R.E.;
RT   "Whole-genome analysis of photosynthetic prokaryotes.";
RL   Science 298:1616-1620(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA   Gerdes S., Kyrpides N., Overbeek R.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY142910; AAN87514.1; -.
FT   NON_TER     545    545
SQ   SEQUENCE   545 AA;  57927 MW;  B9F6DB02DBC18383 CRC64;

Alignment Scores:
Pred. No.:              8.16e-87        Length:         545
Score:                  1804.50         Matches:        360
Percent Similarity:     81.32%          Conservative:   84
Best Local Similarity:  65.93%          Mismatches:     99
Query Match:            60.80%          Indels:         3
DB:                     2               Gaps:           3

US-09-077-574A-1 (1-1647) x Q8GB95 (1-545)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db          2 AlaLysMetThrValPheAsnGluGluAlaArgArgAlaLeuGluLysGlyValAsnThr 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db         22 LeuAlaGluAlaValArgValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db         42 LysPheGlySerProLeuIleThrAsnAspGlyValThrIleAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db         62 GluAsnProIleGluAsnMetGlyAlaGlnLeuValLysGluValAlaThrLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db         82 AspValAlaGlyAspGlyThrThrThrAlaThrIleLeuAlaGlnAlaIleIleArgGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        102 GlyMetLysAsnValAlaAlaGlyAlaAsnProMetValIleLysArgGlyIleGluLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        122 AlaValAlaThrCysValGluGluIleLysAsnLeuSerLysProValGluSerLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        142 AlaIleAlaGlnValAlaSerIleSerAla---AlaAspAlaThrIleGlySerLeuIle 160

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        161 AlaGluAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluSerLysGly 180

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db        181 PheSerThrAspLeuGluValValGluGlyMetAsnPheAspArgGlyTyrIleSerPro 200

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db        201 TyrMetIleThrAspProAspLysMetGluAlaValLeuAsnAspProTyrIleLeuIle 220

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db        221 ThrAspLysLysIleSerAlaIleLysAspIleLeuProValLeuGluArgIleValGln 240

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        241 SerGlyLysGlnLeuMetIleIleSerGluAspIleGluGlyGluAlaLeuAlaThrLeu 260

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db        261 ValValAsnLysLeuArgGlyThrPheThrCysValGlyValLysAlaProGlyPheGly 280

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db        281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyArgValValSer 300

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db        301 GluGluValGlyLeuLysLeuGluAsnAlaThrMetAspMetLeuGlyArgAlaArgGln 320

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db        321 ValArgIleSerLysGluGluThrIleIleValAspGlyAlaGlySerGlnAspAspIle 340

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db        341 LysAlaArgIleAlaGlnIleArgArgGlnTyrGluGluSerThrSerGluPheAspLys 360

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db        361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleGlnValGly 380

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db        381 AlaAlaThrGluThrGluLeuLysGluLysLysLeuArgIleGluAspAlaLeuAsnAla 400

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db        401 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyThrAlaLeuValSerIle 420

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db        421 GlnLysAlaLeuAspGluValGluValProAlaGlyAspGluAlaThrGlyValAlaIle 440

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db        441 IleLysArgSerLeuGluGluProLeuArgGlnIleAlaAsnAsnAlaGlyTyrGluGly 460

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db        461 SerValValValGluLysValLysAsnLeuProLysGlyGlnGlyPheAsnAlaAlaThr 480
```

```
Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
                 |||||||||:::|||   ||||||:::::|||||   ||||||||||||
Db        481 GluValTyrGluAspMetIleAlaAlaGlyIleValAspProAlaLysValThrArgSer 500

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
                 |||||||||||||||||||:::|||::::::|||||||||||||||      :::|||
Db        501 AlaLeuGlnAsnAlaAlaSerIleAlaAlaMetLeuLeuThrThrGluAlaIleValAla 520

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
                 :::|||||||||   ||||||:::   |||         |||||||||   |||||||||
Db        521 AspLysProGlu---LysLysGluAlaProGlyMetProGlyGlyMetProGlyMetGly 539

Qy       1627 GGTATGGAC---GGTATG 1641
                 |||||||||   ||||||
Db        540 GlyMetAspMetGlyMet 545
```

RESULT 6
Q9AJB5
ID Q9AJB5 PRELIMINARY; PRT; 549 AA.
AC Q9AJB5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN GROEL.
OS Pseudoalteromonas sp. PS1M3.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Pseudoalteromonas.
OX NCBI_TaxID=87791;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PS1M3;
RA Kurusu Y., Nakamura T.;
RT "Moleculer chaperone of Psychrophile.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; AB057417; BAB39465.1; -.
DR HSSP; P06139; 1GRL.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 549 AA; 57346 MW; 7AFCC2E8A2F736AF CRC64;

Alignment Scores:

Pred. No.:	2.14e-86	Length:	549
Score:	1796.50	Matches:	360
Percent Similarity:	80.11%	Conservative:	79
Best Local Similarity:	65.69%	Mismatches:	108
Query Match:	60.53%	Indels:	1
DB:	2	Gaps:	1

US-09-077-574A-1 (1-1647) x Q9AJB5 (1-549)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
                |||||:::|||||:::|||||         |||||   |||:::      |||||
Db          1 MetAlaAlaLysGluValLeuPheAlaGlyAspAlaArgAlaLysMetLeuThrGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
                :::   ||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AspLysSerPheGlySerProValIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
                  :::|||   ||||||||||||||||||||||||||||||||||||:::|||
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
                ||||||:::|||   ||||||||||||   |||||||||   :::||||||||||||
Db        101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
                ||||||||:::::|||          ||||||      ::::::   |||      |||
Db        121 AspLysAlaIleIleAlaAlaValAlaGluLeuLysAlaLeuSerValProCysAlaAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                |||   |||||||||||||||||||||||||||||||||||||      ||||||:::
Db        141 ThrLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspLysGluIleGlyAsp 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGA---GGTGTTATCACAGTTGAGGAA 537
                ||||||||||||||||||   |||||||||:::      ||||||||||||||||||||
Db        161 IleIleAlaGluAlaMetGluLysValGlyArgAsnThrGlyValIleThrValGluGlu 180

Qy        538 GCTAAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTAC 597
                  :::   ||||||      ||||||||||||||||||||:::||||||||||||||||
Db        181 GlyGlnSerLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyr 200

Qy        598 CTCTCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTAT 657
                ||||||||||||||:::   |||   ||||||          ||||||||||||||:::
Db        201 LeuSerProTyrPheIleAsnAsnAlaGluLysGlyAlaValGluLeuAspAsnProPhe 220

Qy        658 ATCCTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAA 717
                ||||||      :::||||||:::::::::::::::::::::::||||||   ||||||
Db        221 IleLeuLeuValAspLysLysValSerAsnIleArgGluLeuLeuProThrLeuGluAla 240

Qy        718 GTTGCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTT 777
                |||||||||   ::::::||||||||||||||||||||||:::||||||||||||||||
Db        241 ValAlaLysAlaSerLysProLeuLeuIleIleAlaGluAspLeuGluGlyGluAlaLeu 260

Qy        778 GCAACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCT 837
                |||||||||||||||||   :::||||||   ::::::|||   ||||||||||||||
Db        261 AlaThrLeuValValAsnAsnMetArgGlyIleValLysValSerAlaValLysAlaPro 280

Qy        838 GGTTTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAA 897
                |||||||||:::|||||||||||||||||:::|||||||||:::||||||||||||||
Db        281 GlyPheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaValLeuThrGlyGlyThr 300

Qy        898 GCAATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACA 957
                |||   |||:::   |||::::::|||||      ::::::      |||||||||
Db        301 ValIleSerGluGluIleGlyLeuGluLeuGluLysAlaThrValGluAspLeuGlyThr 320

Qy        958 GCTAAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCA 1017
                |||||||||||||||||   |||:::::|||||||||:::|||||||||||||:::
Db        321 AlaLysArgValValIleThrLysAspAspThrThrIleIleAspGlyAlaGlyGluGlu 340

Qy       1018 GAAGATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGAT 1077
                |||   |||      ||||||   ||||||:::|||||||||||||   :::|||||
Db        341 GluGlyIleThrGlyArgValSerGlnIleLysAlaGlnIleGluGluAlaThrSerAsp 360

Qy       1078 TATGATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATC 1137
                ||||||:::|||||||||||||||||:::|||||||||   |||||||||||||||||
Db        361 TyrAspLysGluLysLeuGlnGluArgMetAlaLysLeuAlaGlyGlyValAlaValIle 380

Qy       1138 CATGTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCT 1197
                ||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db        381 LysValGlyAlaAlaThrGluIleGluMetLysGluLysLysAspArgValGluAspAla 400

Qy       1198 CTAAATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTT 1257
                ||:::|||||||||||||||||||||||||||:::|||||||||||||   |||
Db        401 LeuHisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeu 420
```

```
Qy       1258 GTCCGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGA 1317
Db        421 ValArgAlaAlaSerLysLeuValGluLeuLeuGlyAspAsnGluAspGlnAsnHisGly 440

Qy       1318 CTTAATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGC 1377
Db        441 ValLysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValThrAsnAlaGly 460

Qy       1378 TATGAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAAT 1437
Db        461 AspGluAlaSerValValIleAsnAlaValLysGlyGlyGluGlyAsnTyrGlyTyrAsn 480

Qy       1438 GCTGCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTT 1497
Db        481 AlaAlaThrGlyGluTyrAsnAspMetIleGluMetGlyIleLeuAspProThrLysVal 500

Qy       1498 ACACGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGC 1557
Db        501 ThrArgSerAlaLeuGlnPheAlaGlySerIleAlaGlyLeuMetIleThrThrGluAla 520

Qy       1558 GCTATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGT 1617
Db        521 MetValAlaGluIleProLysGluGluAlaAlaGlyAlaProAspMetGlyGlyMetGly 540

Qy       1618 GGTATGGGTGGTATGGACGGTATG 1641
Db        541 GlyMetGlyGlyMetGlyGlyMet 548
```

RESULT 7
Q8CX48
ID Q8CX48 PRELIMINARY; PRT; 545 AA.
AC Q8CX48;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chaperonin groEL.
GN GROEL OR SO0704.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015516; AAN53782.1; -.
DR TIGR; SO0704; -.
KW Complete proteome.
SQ SEQUENCE 545 AA; 57079 MW; 7591870B017DD521 CRC64;

Alignment Scores:

Pred. No.:	1.47e-85	Length:	545
Score:	1780.50	Matches:	358
Percent Similarity:	79.71%	Conservative:	78
Best Local Similarity:	65.45%	Mismatches:	108
Query Match:	59.99%	Indels:	3
DB:	16	Gaps:	2

US-09-077-574A-1 (1-1647) x Q8CX48 (1-545)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValValPheGlyAsnAspAlaArgValLysMetLeuAlaGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnIleLeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlySerProLeuIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 ThrGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaValIleAlaAlaValAlaGluLeuLysAsnLeuSerGlnAspCysAlaAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyGln 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaThrAlaMetGluLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 GlnAlaLeuGluAsnGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnLysProGluThrGlySerValGluLeuAspHisProPheVal 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuValAspLysLysIleSerAsnIleArgGluLeuLeuProIleLeuGluGlyLeu 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysThrGlyLysProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspValAlaIleLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleAlaGluGluIleGlyLeuGluLeuGluLysAlaThrLeuGluAspLeuGlyThrAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysArgValValIleThrLysAspAsnThrThrIleIleAspGlyAsnGlyGluGlnAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
Db        341 GlnIleGluAlaArgValSerGlnIleLysGlnGlnIleGluGluSerThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
Db        361 AspLysGluLysLeuGlnGluArgMetAlaLysLeuAlaGlyGlyValAlaValIleLys 380
```

```
Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
Db        421 ArgValAlaSerLysIleAlaAspValGluValAlaAsnGluAspGlnLysHisGlyVal 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
Db        441 ValIleAlaLeuArgAlaMetGluAlaProLeuArgGlnIleAlaThrAsnAlaGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440
Db        461 GluAlaSerValValAlaAsnAsnValLysAsnGlySerGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
Db        481 GlyAsnAspThrTyrGlyAspMetLeuGluMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
Db        501 ArgSerAlaLeuGlnPheAlaAlaSerIleAlaGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
Db        521 ValAlaGluLeu------ProLysAlaAspAlaPro---AspMetGlyGlyMetGlyGly 537

Qy       1621 ATGGGTGGTATGGACGGTATG 1641
Db        538 MetGlyGlyMetGlyGlyMet 544

RESULT 8
Q8GBB4
ID   Q8GBB4      PRELIMINARY;      PRT;   546 AA.
AC   Q8GBB4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL.
GN   GROEL.
OS   Azotobacter vinelandii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Azotobacter.
OX   NCBI_TaxID=354;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UWR;
RX   MEDLINE=93323980; PubMed=7687328;
RA   Ferat J.L., Michel F.;
RT   "Group II self-splicing introns in bacteria.";
RL   Nature 364:358-361(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UWR;
RA   Ferat J.-L., Legouar M., Michel F.;
RT   "An active group II intron has invaded the genus Azotobacter and is
RT   inserted within the essential groEL gene.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY057439; AAL25964.1; -.
SQ   SEQUENCE   546 AA;  56875 MW;  8E1F4CC93E42A619 CRC64;

Alignment Scores:
Pred. No.:              1.47e-85      Length:          546
Score:                  1780.50       Matches:         359
Percent Similarity:     79.63%        Conservative:    75
Best Local Similarity:  65.87%        Mismatches:      110
Query Match:            59.99%        Indels:          1

DB:                     2             Gaps:            1

US-09-077-574A-1 (1-1647) x Q8GBB4 (1-546)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysGluValLysPheGlyAspSerAlaArgLysLysMetLeuValGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysAlaThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaLysGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuLysAspLysPheGluAsnMetGlyAlaGlnLeuValLysAspValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 AlaAsnAspGluAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
Db        101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
Db        121 AspLysAlaThrIleAlaIleValAlaGluLeuLysSerLeuAlaLysProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluSerIleGlyAsn 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
Db        161 IleIleAlaGluAlaMetAsnLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
Db        181 SerGlyLeuGluAsnGluLeuSerValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
Db        201 SerProTyrPheIleAsnLysProAspThrMetValAlaGluLeuAspAsnProLeuLeu 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
Db        221 LeuLeuValAspLysLysIleSerAsnIleArgGluLeuLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
Db        241 AlaLysSerGlyArgProLeuLeuIleValAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
Db        261 ThrLeuValValAsnAsnMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
Db        281 PheAsnAspArgArgLysAlaMetLeuGlnValIleAlaIleLeuThrGlyAlaThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
Db        301 IleSerGluLysValGlyLeuSerLeuGluSerAlaThrLeuGluHisLeuGlyLysPro 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
Db        321 LysAlaLeuValLeuAsnLysGluAsnThrThrIleMetHisGlyAlaGlyAlaGlnAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
```

```
Db        341 AspIleGluAlaAlaValAlaGlnIleArgLysGlnIleGluGluThrSerSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db        361 AspArgGluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAlaAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValProGlyGlyGlyValAlaLeuVal 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db        421 ArgAlaLeuGlnAlaIleGluGlyLeuLysGlyAspAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db        441 AlaLeuLeuArgArgAlaValGluAlaProLeuArgGlnIleValAlaAsnAlaGlyAsp 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db        461 GluProSerValValValAspLysValLysGlnGlySerGlyAsnPheGlyPheAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db        481 AlaSerGlyValTyrGlyAspMetIleGluMetGlyIleLeuAspProAlaLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db        501 ArgSerAlaLeuGlnAlaAlaSerSerIleGlyGlyLeuMetIleThrThrGluAlaMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGT 1617

Db        521 ValAlaAspIleValGluAspLysAlaAlaProAlaMetProAspMetGlyGlyMetGly 540

Qy       1618 GGTATGGGTGGTATG 1632

Db        541 GlyMetGlyGlyMet 545

RESULT 9
Q9RC20
ID   Q9RC20      PRELIMINARY;      PRT;   539 AA.
AC   Q9RC20;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Protein Cpn60 (60 kDa chaperonin) (groEL protein).
GN   CPN60.
OS   Bacillus sp. MS.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=96470;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS;
RA   Kondo A., Yohda M.;
RT   "Cpn60/10 from Bacillus strain MS.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AB028452; BAA88110.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   539 AA;  57346 MW;  95626C980E30C46F CRC64;
```

Alignment Scores:

Pred. No.:	6.62e-85	Length:	539
Score:	1768.00	Matches:	354
Percent Similarity:	80.55%	Conservative:	85
Best Local Similarity:	64.95%	Mismatches:	98
Query Match:	59.57%	Indels:	8
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x Q9RC20 (1-539)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66

Db          2 AlaLysGluIleLysPheSerGluGluAlaArgArgAlaMetLeuArgGlyValAspLys 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126

Db         22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186

Db         42 LysPheGlySerProLeuIleThrAsnAspGlyValThrIleAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246

Db         62 GluAspProPheGluAsnMetGlyAlaLysLeuValAlaGluValAlaSerLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306

Db         82 AspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaMetIleArgGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366

Db        102 GlyLeuLysAsnValThrAlaGlyAlaAsnProMetGlyIleArgLysGlyIleGluLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426

Db        122 AlaValAlaValAlaValGluGluLeuLysAlaIleSerLysProIleLysGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486

Db        142 SerIleAlaGlnValAlaAlaIleSerAla---AlaAspGluGluValGlyGlnLeuIle 160

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546

Db        161 AlaGluAlaMetGluArgValGlyAsnAspGlyValIleThrLeuGluGluSerLysGly 180

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606

Db        181 PheThrThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrValSerPro 200

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666

Db        201 TyrMetIleThrAspThrGluLysMetGluAlaValLeuGluAsnProTyrIleLeuIle 220

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726

Db        221 ThrAspLysLysIleSerAsnIleGlnAspIleLeuProIleLeuGluGluGlnValValGln 240

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786

Db        241 GlnGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuAlaThrLeu 260

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846

Db        261 ValValAsnLysLeuArgGlyThrPheThrAlaValAlaValLysAlaProGlyPheGly 280

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
```

```
Db        281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleSer 300

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||   :::|||::::::   :::::::::|||||||||   |||   :::
Db        301 GluGluLeuGlyArgGluLeuLysSerAlaThrIleAlaSerLeuGlyArgAlaSerLys 320

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              ||||||:::   ||||||||||||||||||||||||||||||||   ||||||   |||
Db        321 ValValValThrLysGluAsnThrThrIleValAspGlyAlaGlyAspSerGluArgIle 340

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||||||||:::   |||||||||||||||:::|||||||||:::|||:::::::|||||
Db        341 LysAlaArgIleAsnGlnIleArgAlaGlnLeuGluGluThrThrSerGluPheAspArg 360

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              |||||||||||||||||||||||||||||   ||||||||   ||||||   ||||||
Db        361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValArgValIleLysValGly 380

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              |||||||||||||||||:::||||||:::|||   |||:::|||||||||||||||:::
Db        381 AlaAlaThrGluThrGluLeuLysGluArgLysLeuArgIleGluAspAlaLeuAsnSer 400

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||||||||||||||||||||||||||||   ||||||||||||||   :::
Db        401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuMetAsnVal 420

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                    :::      |||:::   |||:::   ||||||      |||:::   |||
Db        421 TyrAsnLysValAlaAlaIleGlu---AlaGluGlyAspGluAlaThrGlyValLysIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              :::   |||::::::|||||||||:::|||||||||||   |||||||||   ||||||
Db        440 ValLeuArgAlaIleGluGluProValArgGlnIleAlaGlnAsnAlaGlyLeuGluGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
              |||::::::|||||:::::::::      |||   |||   ||||||||||||||:::
Db        460 SerValIleValGluArgLeuLysThrGluLysProGlyIleGlyPheAsnAlaAlaThr 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              ||||||:::   |||:::|||:::||||||::::::||||||   ||||||||||||
Db        480 GlyGluTrpValAspMetIleGluAlaGlyIleValAspProThrLysValThrArgSer 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              |||||||||||||||||||||||||||||::::::   ||||||||||||      :::|||
Db        500 AlaLeuGlnAsnAlaAlaSerValAlaAlaMetPheLeuThrThrGluAlaValValAla 519

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
              :::|||||||||      |||                ||||||      ||||||
Db        520 AspLysProGluGluAsnLys----------------GlyGlyAsnAlaGlyMetPro 533

Qy       1627 GGTATGGACGGTATG 1641
                 |||   ||||||
Db        534 AspMetGlyGlyMet 538

RESULT 10
O31198
ID   O31198      PRELIMINARY;      PRT;   546 AA.
AC   O31198;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Heat shock protein 58 (60 kDa chaperonin) (Protein Cpn60) (groEL
DE   protein).
GN   HSP58.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HY-1;
RA   Kim M.J., Ahn B.Y.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF032910; AAB86965.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   546 AA;  57969 MW;  15BC9EC0A8E2A6E3 CRC64;

Alignment Scores:
Pred. No.:              6.63e-85        Length:         546
Score:                  1768.00         Matches:        352
Percent Similarity:     80.07%          Conservative:   86
Best Local Similarity:  64.35%          Mismatches:     101
Query Match:            59.57%          Indels:         8
DB:                     2               Gaps:           3

US-09-077-574A-1 (1-1647) x O31198 (1-546)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::|||:::|||   ::::::      ||||||   ||||||      ||||||:::|||
Db          2 AlaLysAspIleGluTyrAsnGluThrAlaArgArgLysLeuLeuGluGlyValAsnLys 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db         22 LeuAlaAsnAlaValLysValThrLeuGlyProLysGlyArgAsnValValIleAspLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              ||||||:::|||   |||||||||||||||||:::||||||||||||||||||||||
Db         42 LysPheGlyAlaProThrIleThrLysAspGlyValThrValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
              ||||||      |||||||||||||||||||||||||||||:::   ||||||:::
Db         62 GluAspProLeuGluAsnMetGlyAlaGlnMetValLysGluValSerThrLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||:::|||||||||||||||||||||||||||:::|||||||:::|||      |||
Db         82 AspValAlaGlyAspGlyThrThrThrAlaThrIleLeuAlaGlnSerIleIleAsnGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||:::|||   |||   ||||||   ||||||||:::::|||||||||||||||||||
Db        102 GlyLeuLysAsnValThrAlaGlyAlaAsnProMetSerLeuLysArgGlyIleAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              ||||||   |||      :::   :::                        ::::::|||
Db        122 AlaValThrAlaAlaValGluSerIleGlnLysArgAlaValLysIleGluAsnLysLys 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              :::||||||   |||   :::||||||||||||:::|||   |||||||||||:::|||
Db        142 AspIleAlaAsnValAlaSerIleSerAlaAsnAsnAspAsnThrIleGlyAsnLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||   ||||||||||||   ||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetAspLysValGlyLysAspGlyValIleThrValGluGluAlaLysSer 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||||||||||||||||||||||||||||:::|||||||||||||||:::||||||
Db        182 IleGluThrThrLeuAspValValGluGlyMetGlnPheAspArgGlyTyrIleSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              |||   ||||||:::   |||   ||||||      |||:::::|||:::||||||
Db        202 TyrMetValThrAspAlaGluSerMetValAlaThrLeuAsnAspProPheIleLeuIle 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
              :::|||||||||:::|||||||||||:::::   |||||||||:::|||||||:::
Db        222 TyrAspLysLysIleSerSerMetLysAspLeuIleHisIleLeuGluLysValAlaGln 241
```

```
Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db        242 AlaGlyLysProLeuValIleIleSerGluGluValGluGlyGluAlaLeuAlaThrIle 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db        262 ValValAsnThrLeuArgLysThrIleSerCysValAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db        282 AspArgArgLysSerMetLeuGluAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db        302 GluAspLeuGlyMetLysLeuGluAsnThrThrLeuGlnMetLeuGlyArgAlaAsnLys 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db        322 ValThrValAspLysGluAsnThrThrIleIleGluGlyLysGlyGlnThrLysGluIle 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db        342 GlnGlyArgIleGlyGlnIleLysLysGlnIleGluAspThrThrSerGluTyrAspArg 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleHisValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db        382 AlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db        402 ThrArgAlaAlaValGluGluGlyIleValProGlyGlyGlyLeuThrLeuLeuLysAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db        422 GlnGluAlaValGlySerLeuLys---LeuAspGlyAspGluAlaThrGlyAlaLysIle 440

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db        441 IlePheArgAlaLeuGluGluProIleArgMetIleThrSerAsnAlaGlyLeuGluGly 460

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db        461 SerValIleValGluHisAlaLysAlaLysLysGlyAsnGluGlyPheAsnAlaLeuThr 480

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db        481 MetValTrpGluAspMetIleGlnAlaGlyValValAspProAlaLysValValArgSer 500

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db        501 AlaLeuGlnAsnAlaAlaSerIleGlySerMetIleLeuThrThrGluValThrIleThr 520

Qy       1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCT----------GGCGGTGGT 1611
Db        521 AspLysProAsp------LysAspAlaProAsnProMetAlaGlyMetGlyGlyGlyGly 538

Qy       1612 ATGGGTGGTATGGGTGGTATG 1632
Db        539 MetGlyGlyMetGlyGlyMet 545
```

RESULT 11
Q9EZV4
ID Q9EZV4 PRELIMINARY; PRT; 539 AA.
AC Q9EZV4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN GROEL.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31783;
RA Snapyan M., Gochikyan A., Weigel P., Sakanyan V.;
RT "Effect of thermostable chaperonins on synthesis of proteins in
RT vitro.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; AF274871; AAG44815.1; -.
DR HSSP; P06139; 1GRL.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 539 AA; 57293 MW; 2F51B6FFBB887428 CRC64;

Alignment Scores:

Pred. No.:	3.57e-84	Length:	539
Score:	1754.00	Matches:	352
Percent Similarity:	80.55%	Conservative:	87
Best Local Similarity:	64.59%	Mismatches:	98
Query Match:	59.10%	Indels:	8
DB:	2	Gaps:	3

US-09-077-574A-1 (1-1647) x Q9EZV4 (1-539)

```
Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
Db          2 AlaLysGluIleLysPheSerGluGluAlaArgArgAlaMetLeuArgGlyValAspLys 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
Db         22 LeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeuGluLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
Db         42 LysPheGlySerProLeuIleThrAsnAspGlyValThrIleAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
Db         62 GluAspProPheGluAsnMetGlyAlaLysLeuValValGluValAlaSerLysThrAsn 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
Db         82 AspValAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaMetIleArgGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
Db        102 GlyLeuLysAsnValThrAlaGlyAlaAsnProMetGlyIleArgLysGlyIleGluLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
Db        122 AlaValAlaValAlaValGluGluLeuLysAlaIleSerLysProIleGlnGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
Db        142 SerIleAlaGlnValAlaAlaIleSerAla---AlaAspGluGluValGlyGlnLeuIle 160

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
Db        161 AlaGluAlaMetGluArgValGlyAsnAspGlyValIleThrLeuGluGluSerLysGly 180
```

```
Qy       547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
Db       181 PheThrThrGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrAlaSerPro 200

Qy       607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
Db       201 TyrMetIleThrAspThrGluLysMetGluAlaValLeuGluAsnProTyrIleLeuIle 220

Qy       667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
Db       221 ThrAspLysLysIleSerAsnIleGlnAspLeuLeuProIleLeuGluGlnValValGln 240

Qy       727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
Db       241 GlnGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAlaThrLeu 260

Qy       787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
Db       261 ValValAsnLysLeuArgGlyThrPheThrAlaValAlaValLysAlaProGlyPheGly 280

Qy       847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
Db       281 AspArgArgLysAlaMetLeuGluAspIleAlaIleLeuThrGlyGlyGluValIleSer 300

Qy       907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
Db       301 GluGluLeuGlyArgGluLeuLysSerThrThrIleAlaSerLeuGlyArgAlaSerLys 320

Qy       967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
Db       321 ValValValThrLysGluAsnThrThrIleValGluGlyAlaGlyAspSerAspArgIle 340

Qy      1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
Db       341 LysAlaArgIleAsnGlnIleArgAlaGlnLeuGluGluThrThrSerGluPheAspArg 360

Qy      1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
Db       361 GluLysLeuGlnGluArgLeuAlaLysLeuAlaGlyGlyValAlaValIleLysValGly 380

Qy      1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
Db       381 AlaAlaThrGluThrGluLeuLysGluArgLysLeuArgIleGluAspAlaLeuAspSer 400

Qy      1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
Db       401 ThrArgAlaAlaValGluGluGlyIleValAlaGlyGlyGlyThrAlaLeuMetAsnVal 420

Qy      1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
Db       421 TyrSerLysValAlaAlaIleGlu---AlaGluGlyAspGluAlaThrGlyValLysIle 439

Qy      1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
Db       440 ValLeuArgAlaIleGluGluProValArgGlnIleAlaGlnAsnAlaGlyLeuGluGly 459

Qy      1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
Db       460 SerValIleValGluArgLeuLysThrGluLysProGlyIleGlyPheAsnAlaAlaThr 479

Qy      1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
Db       480 GlyGluTrpValAsnMetIleGluAlaGlyIleValAspProThrLysValThrArgSer 499

Qy      1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
Db       500 AlaLeuGlnAsnAlaAlaSerValAlaAlaMetPheLeuThrThrGluAlaValValAla 519

Qy      1567 GAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGTATGGGT 1626
Db       520 AspLysProGluGluAsnLys----------------GlyGlyAsnAsnGlyMetPro 533

Qy      1627 GGTATGGACGGTATG 1641
Db       534 AspMetGlyGlyMet 538
```

RESULT 12
Q93FU8
ID Q93FU8 PRELIMINARY; PRT; 545 AA.
AC Q93FU8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE groEL-like protein (60 kDa chaperonin) (Protein Cpn60) (groEL
DE protein).
GN GROEL.
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=548;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshida N., Oeda K., Watanabe E., Mikami T., Fukita Y., Nishimura K.,
RA Komai K., Matsuda K.;
RT "A GroEL homolog produced by endosymbionts in antlions as a paralytic
RT toxin against insects.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; AF306521; AAL09389.1; -.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 545 AA; 57023 MW; DB3B7A017DEDA85F CRC64;

Alignment Scores:

Pred. No.:	1.52e-83	Length:	545
Score:	1742.00	Matches:	353
Percent Similarity:	78.02%	Conservative:	73
Best Local Similarity:	64.65%	Mismatches:	116
Query Match:	58.69%	Indels:	4
DB:	2	Gaps:	2

US-09-077-574A-1 (1-1647) x Q93FU8 (1-545)

```
Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60
Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
Db         21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy        121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
Db         41 AspLysSerPheGlyAlaProThrIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy        181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
Db         61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy        241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
Db         81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnAlaIleVal 100

Qy        301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
```

```
Db        101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy        361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420

Db        121 AspLysAlaValValAlaAlaValGluGluLeuLysAlaLeuSerValProCysSerAsp 140

Qy        421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480

Db        141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyLys 160

Qy        481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540

Db        161 LeuIleAlaGluAlaMetAspLysValGlyLysGluGlyValIleThrValGluAspGly 180

Qy        541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600

Db        181 ThrGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy        601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660

Db        201 SerProTyrPheIleAsnLysProGluThrGlyAlaValGluLeuGluSerProPheIle 220

Qy        661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720

Db        221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy        721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780

Db        241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy        781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840

Db        261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy        841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900

Db        281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaThrLeuThrGlyGlyThrVal 300

Qy        901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960

Db        301 IleSerGluGluIleGlyMetGluLeuGluLysAlaThrLeuGluAspLeuGlyGlnAla 320

Qy        961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020

Db        321 LysArgValValIleAsnLysAspThrThrThrIleIleAspGlyValGlyAspGluAla 340

Qy       1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080

Db        341 AlaIleGlnGlyArgValThrGlnIleArgGlnGlnIleGluGluAlaThrSerAspTyr 360

Qy       1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140

Db        361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy       1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200

Db        381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy       1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260

Db        401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy       1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320

Db        421 ArgValAlaSerLysIleAlaGlyLeuLysGlyGlnAsnGluAspGlnAsnValGlyIle 440

Qy       1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380

Db        441 LysValAlaLeuArgAlaMetGluAlaProLeuArgGlnIleValLeuAsnCysGlyGlu 460

Qy       1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCT 1440

Db        461 GluProSerValValAlaAsnThrValLysAlaGlyAspGlyAsnTyrGlyTyrAsnAla 480

Qy       1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500

Db        481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy       1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560

Db        501 ArgSerAlaLeuGlnTyrAlaAlaSerValAlaGlyLeuMetIleThrThrGluCysMet 520

Qy       1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGC------GGTGGTATG 1614

Db        521 ValThrAspLeu------ProLysSerAspAlaProAspLeuGlyAlaAlaGlyGlyMet 538

Qy       1615 GGTGGTATGGGTGGTATG 1632

Db        539 GlyGlyMetGlyGlyMet 544

RESULT 13
O33688
ID   O33688      PRELIMINARY;      PRT;   538 AA.
AC   O33688;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60) (Fragment).
GN   GROEL.
OS   primary endosymbiont of Sitophilus oryzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae.
OX   NCBI_TaxID=2342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SFr;
RX   MEDLINE=98042468; PubMed=9367844;
RA   Charles H., Heddi A., Guillaud J., Nardon C., Nardon P.;
RT   "A molecular aspect of symbiotic interactions between the weevil
RT   Sitophilus oryzae and its endosymbiotic bacteria: over-expression of a
RT   chaperonin.";
RL   Biochem. Biophys. Res. Commun. 239:769-774(1997).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF005236; AAB97670.1; -.
DR   HSSP; P06139; 1JON.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
FT   NON_TER     538     538
SQ   SEQUENCE   538 AA;  56683 MW;  EFD17668C0F09D15 CRC64;

Alignment Scores:
Pred. No.:              3.12e-83        Length:         538
Score:                  1736.00         Matches:        348
Percent Similarity:     78.86%          Conservative:   81
Best Local Similarity:  63.97%          Mismatches:     109
Query Match:            58.49%          Indels:         6
DB:                     2               Gaps:           2

US-09-077-574A-1 (1-1647) x O33688 (1-538)

Qy          1 ATGGCTTCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTA 60

Db          1 MetAlaAlaLysAspValLysPheGlyAsnAspAlaArgValLysMetLeuArgGlyVal 20

Qy         61 GATAAACTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATT 120
```

```
Db        21 AsnValLeuAlaAspAlaValLysValThrLeuGlyProLysGlyArgAsnValValLeu 40

Qy       121 GAAAAGTCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATT 180
             :::|||||||||||||:::|||||||||||||||||||||||||||||||||||:::||||||
Db        41 AspLysSerPheGlyAlaProValIleThrLysAspGlyValSerValAlaArgGluIle 60

Qy       181 GAACTTGAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        61 GluLeuGluAspLysPheGluAsnMetGlyAlaGlnMetValLysGluValAlaSerLys 80

Qy       241 ACTAGCGATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTAT 300
               :::|||   |||||||||||||||||||||||||||||||||||||||:::|||
Db        81 AlaAsnAspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaGlnSerIleVal 100

Qy       301 CGTGAAGGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATA 360
                ||||||:::|||   ||||||||||||   |||||||||   :::||||||||||||
Db       101 AsnGluGlyLeuLysAlaValAlaAlaGlyMetAsnProMetAspLeuLysArgGlyIle 120

Qy       361 GATAAAGCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGAC 420
             ||||||||||||:::|||      :::||||||      ::::::   |||       |||
Db       121 AspLysAlaValIleAlaAlaValGluGluLeuLysLysLeuSerValProCysSerAsp 140

Qy       421 CAAAAAGAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAAT 480
                |||   |||||||||||||||||||||||||||||||||||||||   ||:::|||
Db       141 SerLysAlaIleAlaGlnValGlyThrIleSerAlaAsnSerAspGluThrValGlyThr 160

Qy       481 ATCATAGCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCT 540
             :::|||||||||||||||||||||||||||||   ||||||||||||||||||||
Db       161 LeuIleAlaGluAlaMetAlaLysValGlyLysGluGlyValIleThrValGluGluGly 180

Qy       541 AAAGGTCTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTC 600
             ||||||:::      |||||||||||||||||||:::|||||||||||||||||||||
Db       181 SerGlyLeuGlnAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeu 200

Qy       601 TCTCCATACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATC 660
             |||||||||||||||      ||||||                ||||||::::::|||:::|||
Db       201 SerProTyrPheValAsnLysProGluThrGlyAlaIleGluLeuGluSerProPheIle 220

Qy       661 CTTTGTAATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTT 720
             |||      :::|||||||||:::::::::::::::||||||||:::||||||   |||
Db       221 LeuLeuAlaAspLysLysIleSerAsnIleArgGluMetLeuProValLeuGluAlaVal 240

Qy       721 GCTAAAGTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCA 780
             ||||||      :::||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AlaLysAlaGlyLysProLeuLeuIleIleAlaGluAspValGluGlyGluAlaLeuAla 260

Qy       781 ACACTTGTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGT 840
             |||||||||||||||   :::||||||   ::::::|||   |||||||||||||||||
Db       261 ThrLeuValValAsnThrMetArgGlyIleValLysValAlaAlaValLysAlaProGly 280

Qy       841 TTTGGTGAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCA 900
             ||||||:::|||||||||||||||||:::||||||||||||||||||   |||
Db       281 PheGlyAspArgArgLysAlaMetLeuGlnAspIleAlaIleLeuThrAlaGlyThrVal 300

Qy       901 ATATTTGAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCT 960
             |||   |||:::   |||::::::||||||      :::|||      :::|||   |||
Db       301 IleSerGluGluIleGlyLeuGluLeuGluLysAlaThrLeuGluAspMetGlyGlnAla 320

Qy       961 AAACGTGTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAA 1020
             |||||||||||||||   |||:::   ||||||||:::||||||   |||
Db       321 LysArgValValIleThrLysAspThrThrThrIleIleAspGlyValGlyAspLysAla 340

Qy      1021 GATATTAAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTAT 1080
                |||   :::||||||   ||||||      |||   :::|||   :::||||||||||
Db       341 LeuIleAspSerArgValThrGlnIleAsnGlnGlnArgAspGluAlaThrSerAspTyr 360

Qy      1081 GATCGTGAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCAT 1140
             |||||||||||||||||||||||:::|||||||||   ||||||||||||||||||
Db       361 AspArgGluLysLeuGlnGluArgValAlaLysLeuAlaGlyGlyValAlaValIleLys 380

Qy      1141 GTTGGAGCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTA 1200
             ||||||||||||||||   ||||||||||||||||   ||||||||||||||||||
Db       381 ValGlyAlaAlaThrGluValGluMetLysGluLysLysAlaArgValGluAspAlaLeu 400

Qy      1201 AATGCAACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTC 1260
             :::|||||||||||||||||||||||||||:::|||   |||||||||   |||   :::
Db       401 HisAlaThrArgAlaAlaValGluGluGlyValValAlaGlyGlyGlyValAlaLeuIle 420

Qy      1261 CGCTCCATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTT 1320
             |||              :::   :::::::::      ::::::|||:::      |||:::
Db       421 ArgValAlaAsnSerIleAlaGluLeuArgGlyAspAsnGluAspGlnAsnValGlyIle 440

Qy      1321 AATATCATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTAT 1380
                :::   ||||||::::::|||   ||||||||||||||   ||||||||||||
Db       441 LysValAlaArgArgAlaMetGluAlaProLeuArgGlnIleValAlaAsnAlaGlyGlu 460

Qy      1381 GAAGGTTCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAAGATGGTTTTGGATTTAATGCT 1440
             |||   |||::::::      ||||||:::      :::          |||:::||||||
Db       461 GluProSerValIleAlaAsnLysValLysAlaGlyGluGlyAsnThrGlyTyrAsnAla 480

Qy      1441 GCATCAGGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACA 1500
             |||:::   ||||||   ::::::|||      |||::::::||||||   |||||||||
Db       481 AlaThrGluGluTyrGlyAsnMetIleAspMetGlyIleLeuAspProThrLysValThr 500

Qy      1501 CGTATTGCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCT 1560
             |||   |||||||||   ||||||||:::|||   |||:::::|||||||||||||
Db       501 ArgSerAlaLeuGlnTyrAlaAlaSerIleAlaGlyLeuMetIleThrThrGluCysMet 520

Qy      1561 ATTGCTGAAAAACCAGAACCTAAAAAAGATATGCCTATGCCTGGCGGTGGTATGGGTGGT 1620
             :::   :::      :::||||||:::|||   |||                :::||||||
Db       521 ValThrAsp------GlnProLysGluAspLysPro------------AspLeuGlyGly 534

Qy      1621 ATGGGTGGTATG 1632
             ||||||||||||
Db       535 MetGlyGlyMet 538
```

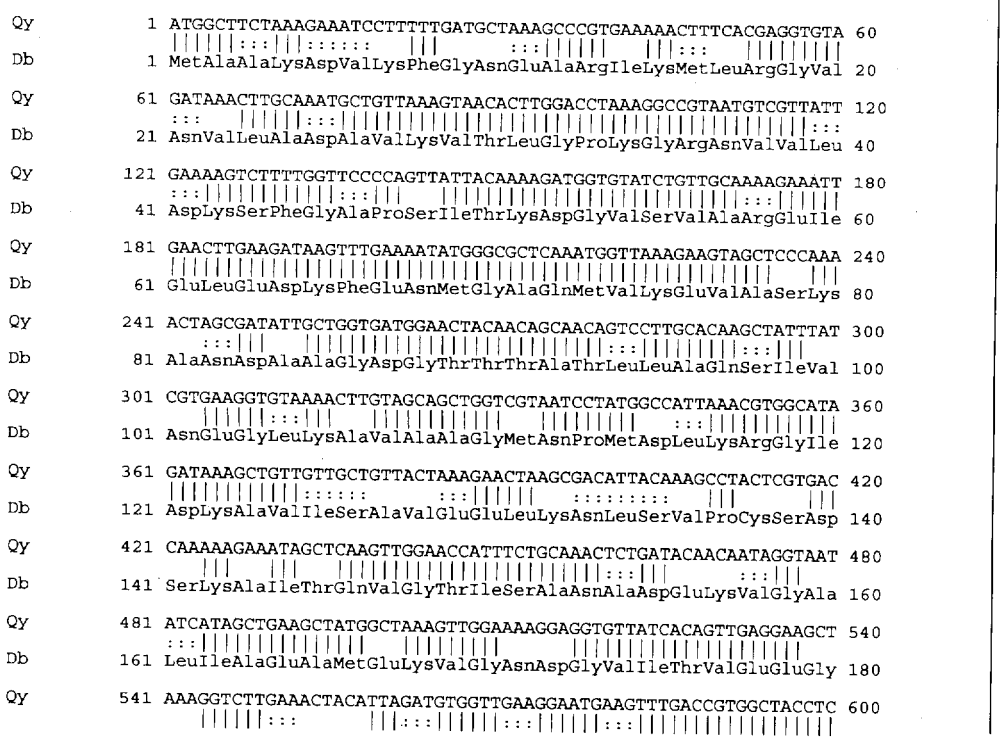
RESULT 14
Q8RNU2
ID Q8RNU2 PRELIMINARY; PRT; 546 AA.
AC Q8RNU2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Heat shock protein B subunit (60 kDa chaperonin) (Protein Cpn60)
DE (groEL protein).
GN HSPB.
OS Helicobacter pylori.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-CTX1;
RX MEDLINE=20123532; PubMed=10660136;
RA Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
RA Garcia-de la Guarda R., Urra S., Venegas A.;
RT "Serological response to Helicobacter pylori recombinant antigens in
RT Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
RT gastric cancer.";
RL APMIS 107:1069-1078(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-CTX1;
RA Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
RA Bruce E., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;
RT "Cloning and comparison of ten gene sequences of a Chilean
RT Helicobacter pylori strain with other Helicobacter pylori strains
RT revealed higher amino acid sequence variability for VacA and CagA
RT virulence factors.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC   PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC   CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC   7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

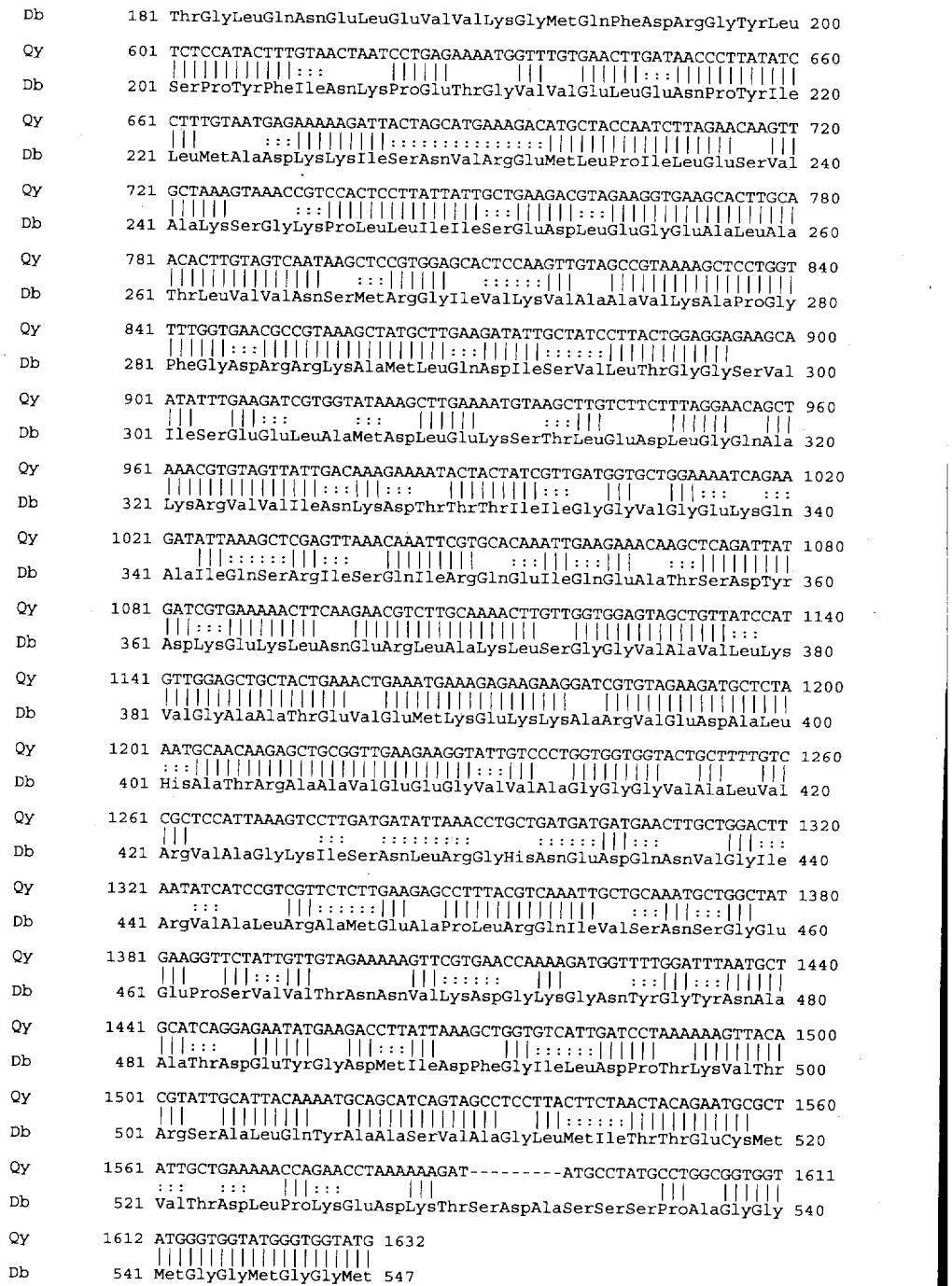
```
DR   EMBL; AF479030; AAL86900.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   546 AA;  58242 MW;  B3DAE13619F26B32 CRC64;

Alignment Scores:
Pred. No.:              1.59e-82      Length:         546
Score:                  1722.50       Matches:        350
Percent Similarity:     77.47%        Conservative:   73
Best Local Similarity:  64.10%        Mismatches:     120
Query Match:            58.04%        Indels:         3
DB:                     2             Gaps:           2

US-09-077-574A-1 (1-1647) x Q8RNU2 (1-546)

Qy          7 TCTAAAGAAATCCTTTTTGATGCTAAAGCCCGTGAAAAACTTTCACGAGGTGTAGATAAA 66
              :::||||||||   |||        ||||||      |||      ||||||   :::
Db          2 AlaLysGluIleLysPheSerAspSerAlaArgAsnLeuLeuPheGluGlyValArgGln 21

Qy         67 CTTGCAAATGCTGTTAAAGTAACACTTGGACCTAAAGGCCGTAATGTCGTTATTGAAAAG 126
              |||   :::|||||||||||||||:::|||||:::||||||||||:::|||:::|||
Db         22 LeuHisAspAlaValLysValThrMetGlyProArgGlyArgAsnValLeuIleGlnLys 41

Qy        127 TCTTTTGGTTCCCCAGTTATTACAAAAGATGGTGTATCTGTTGCAAAAGAAATTGAACTT 186
              |||:::|||:::|||   |||||||||||||||||||||||||||||||||||||||||
Db         42 SerTyrGlyAlaProSerIleThrLysAspGlyValSerValAlaLysGluIleGluLeu 61

Qy        187 GAAGATAAGTTTGAAAATATGGGCGCTCAAATGGTTAAAGAAGTAGCTCCCAAAACTAGC 246
                              |||||||||||||||:::|||||||||||||   ||||||:::
Db         62 SerCysProValAlaAsnMetGlyAlaGlnLeuValLysGluValAlaSerLysThrAla 81

Qy        247 GATATTGCTGGTGATGGAACTACAACAGCAACAGTCCTTGCACAAGCTATTTATCGTGAA 306
              |||   |||||||||||||||||||||||||||||||||   :::|||::::::|||
Db         82 AspAlaAlaGlyAspGlyThrThrThrAlaThrValLeuAlaTyrSerIlePheLysGlu 101

Qy        307 GGTGTAAAACTTGTAGCAGCTGGTCGTAATCCTATGGCCATTAAACGTGGCATAGATAAA 366
              |||::::::   :::   ||||||   ||||||:::   :::|||||||||:::||||||
Db        102 GlyLeuArgAsnIleThrAlaGlyAlaAsnProIleGluValLysArgGlyMetAspLys 121

Qy        367 GCTGTTGTTGCTGTTACTAAAGAACTAAGCGACATTACAAAGCCTACTCGTGACCAAAAA 426
              |||      |||:::      ||||||         :::|||                ::::::
Db        122 AlaAlaGluAlaIleIleAsnGluLeuLysLysAlaSerLysLysValGlyGlyLysGlu 141

Qy        427 GAAATAGCTCAAGTTGGAACCATTTCTGCAAACTCTGATACAACAATAGGTAATATCATA 486
              ||||||   ||||||   |||||||||||||||||||||      |||||   :::|||
Db        142 GluIleThrGlnValAlaThrIleSerAlaAsnSerAspHisAsnIleGlyLysLeuIle 161

Qy        487 GCTGAAGCTATGGCTAAAGTTGGAAAAGGAGGTGTTATCACAGTTGAGGAAGCTAAAGGT 546
              |||:::||||||   ||||||||||||   ||||||||||||||||||||||||||||||
Db        162 AlaAspAlaMetGluLysValGlyLysAspGlyValIleThrValGluGluAlaLysGly 181

Qy        547 CTTGAAACTACATTAGATGTGGTTGAAGGAATGAAGTTTGACCGTGGCTACCTCTCTCCA 606
              :::|||      ||||||||||||||||||:::|||||||||||||||||||||||||||
Db        182 IleGluAspGluLeuAspValValGluGlyMetGlnPheAspArgGlyTyrLeuSerPro 201

Qy        607 TACTTTGTAACTAATCCTGAGAAAATGGTTTGTGAACTTGATAACCCTTATATCCTTTGT 666
              ||||||||||||||   ||||||||      :::|||||||||   |||||||||
Db        202 TyrPheValThrAsnAlaGluLysMetThrAlaGlnLeuAspAsnAlaTyrIleLeuLeu 221

Qy        667 AATGAGAAAAAGATTACTAGCATGAAAGACATGCTACCAATCTTAGAACAAGTTGCTAAA 726
                 :::|||||||||:::||||||||||||:::||||||:::||||||:::      |||
Db        222 ThrAspLysLysIleSerSerMetLysAspIleLeuProLeuLeuGluLysThrMetLys 241

Qy        727 GTAAACCGTCCACTCCTTATTATTGCTGAAGACGTAGAAGGTGAAGCACTTGCAACACTT 786
                    :::|||||||||||||||||||||||:::|||||||||||||||   ||||||
Db        242 GluGlyLysProLeuLeuIleIleAlaGluAspIleGluGlyGluAlaLeuThrThrLeu 261

Qy        787 GTAGTCAATAAGCTCCGTGGAGCACTCCAAGTTGTAGCCGTAAAAGCTCCTGGTTTTGGT 846
              ||||||||||||||||||||   |||   :::   ||||||||||||||||||||||||
Db        262 ValValAsnLysLeuArgGlyValLeuAsnIleAlaAlaValLysAlaProGlyPheGly 281

Qy        847 GAACGCCGTAAAGCTATGCTTGAAGATATTGCTATCCTTACTGGAGGAGAAGCAATATTT 906
              :::|||||||||   ||||||:::|||||||||||||||||||||||:::   |||
Db        282 AspArgArgLysGluMetLeuLysAspIleAlaIleLeuThrGlyGlyGlnValIleSer 301

Qy        907 GAAGATCGTGGTATAAAGCTTGAAAATGTAAGCTTGTCTTCTTTAGGAACAGCTAAACGT 966
              |||:::   |||:::   |||||||||       :::      ||||||   |||   |||
Db        302 GluGluLeuGlyLeuSerLeuGluAsnAlaGluValGluPheLeuGlyLysAlaGlyArg 321

Qy        967 GTAGTTATTGACAAAGAAAATACTACTATCGTTGATGGTGCTGGAAAATCAGAAGATATT 1026
              :::|||||||||||:::|||||||||||||||||||||   |||   |||:::|||:::
Db        322 IleValIleAspLysAspAsnThrThrIleValAspGlyLysGlyHisSerAspAspVal 341

Qy       1027 AAAGCTCGAGTTAAACAAATTCGTGCACAAATTGAAGAAACAAGCTCAGATTATGATCGT 1086
              |||   ||||||   ||||||:::   ||||||      |||:::|||||||||||:::
Db        342 LysAspArgValAlaGlnIleLysThrGlnIleAlaSerThrThrSerAspTyrAspLys 361

Qy       1087 GAAAAACTTCAAGAACGTCTTGCAAAACTTGTTGGTGGAGTAGCTGTTATCCATGTTGGA 1146
              |||||||||||||||||||||||||||||   |||||||||||||||||   ||||||
Db        362 GluLysLeuGlnGluArgLeuAlaLysLeuSerGlyGlyValAlaValIleLysValGly 381

Qy       1147 GCTGCTACTGAAACTGAAATGAAAGAGAAGAAGGATCGTGTAGAAGATGCTCTAAATGCA 1206
              ||||||:::|||   |||||||||||||||||||||||||:::|||||||||:::|||
Db        382 AlaAlaSerGluValGluMetLysGluLysLysAspArgValAspAspAlaLeuSerAla 401

Qy       1207 ACAAGAGCTGCGGTTGAAGAAGGTATTGTCCCTGGTGGTGGTACTGCTTTTGTCCGCTCC 1266
              |||:::|||||||||||||||||||||||   |||||||||   |||   :::|||:::
Db        402 ThrLysAlaAlaValGluGluGlyIleValIleGlyGlyGlyAlaAlaLeuIleArgAla 421

Qy       1267 ATTAAAGTCCTTGATGATATTAAACCTGCTGATGATGATGAACTTGCTGGACTTAATATC 1326
                 :::   :::                        |||||||||      |||      |||
Db        422 AlaGlnLysValHis------LeuAsnLeuHisAspAspGluLysValGlyTyrGluIle 439

Qy       1327 ATCCGTCGTTCTCTTGAAGAGCCTTTACGTCAAATTGCTGCAAATGCTGGCTATGAAGGT 1386
              |||   |||:::::::::   ||||||   |||||||||   ||||||||||||:::|||
Db        440 IleMetArgAlaIleLysAlaProLeuAlaGlnIleAlaIleAsnAlaGlyTyrAspGly 459

Qy       1387 TCTATTGTTGTAGAAAAAGTTCGTGAACCAAAAGATGGTTTTGGATTTAATGCTGCATCA 1446
                 ::::||||||   :::|||   :::   :::      ||||||||||||||:::::
Db        460 GlyValValValAsnGluValGluLysHisGluGlyHisPheGlyPheAsnAlaSerAsn 479

Qy       1447 GGAGAATATGAAGACCTTATTAAAGCTGGTGTCATTGATCCTAAAAAAGTTACACGTATT 1506
              |||:::|||   |||:::   |||   |||:::|||||||||   ||||||   ||||||
Db        480 GlyLysTyrValAspMetPheLysGluGlyIleIleAspProLeuLysValGluArgIle 499

Qy       1507 GCATTACAAAATGCAGCATCAGTAGCCTCCTTACTTCTAACTACAGAATGCGCTATTGCT 1566
              ||||||||||||||   ||||||:::|||||||||||||||||||||||      :::
Db        500 AlaLeuGlnAsnAlaValSerValSerSerLeuLeuLeuThrThrGluAlaThrValHis 519

Qy       1567 GAAAAACCAGAACCTAAAAAAAGATATGCCTATGCCTGGC---GGTGGTATGGGTGGTATG 1623
              |||      |||   |||            ||||||      ||||||||||||||||||
Db        520 GluIleLysGluGluLysAlaAlaProAlaMetProAspMetGlyGlyMetGlyGlyMet 539

Qy       1624 GGTGGTATGGACGGTATG 1641
              |||||||||   ||||||
Db        540 GlyGlyMetGlyGlyMet 545

RESULT 15
Q8KIW5
ID   Q8KIW5      PRELIMINARY;      PRT;   548 AA.
AC   Q8KIW5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN   GROEL.
OS   Buchnera sp. 150.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=168690;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=150 Neuville-1;
```

```
RA   Delmotte F., Sabater B., Simon J.C., Latorre A.;
RT   "Coevolution between mitochondrialand simbiotic DNA at intraspecific
RT   level.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=150 Neuville-1;
RA   Delmotte F.;
RL   Thesis (2001), University of Valencia, Spain.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AJ344978; CAC86959.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   548 AA;  57853 MW;  659886423C19529D CRC64;
```

Alignment Scores:			
Pred. No.:	4.17e-82	Length:	548
Score:	1714.50	Matches:	338
Percent Similarity:	78.61%	Conservative:	92
Best Local Similarity:	61.79%	Mismatches:	114
Query Match:	57.77%	Indels:	3
DB:	2	Gaps:	1

US-09-077-574A-1 (1-1647) x Q8KIW5 (1-548)





Search completed: January 28, 2004, 13:28:47
Job time : 98 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:10:00 ; Search time 47 Seconds
(without alignments)
1850.684 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2708	100.0	548	18	AAW16678	Lawsonia intracell
2	1864.5	68.9	545	20	AAY23930	Consensus mino aci
3	1848	68.2	549	20	AAY23909	Amino acid sequenc
4	1825.5	67.4	547	22	AAB69060	Pseudomonas aerugi
5	1820.5	67.2	547	16	AAR67381	L. pneumophila Htp
6	1819	67.2	574	22	AAB69061	pTrcHisB expressio
7	1814.5	67.0	545	20	AAY23915	Amino acid sequenc
8	1812.5	66.9	544	21	AAY75747	Neisseria meningit
9	1805.5	66.7	544	24	ABP77245	N. gonorrhoeae ami
10	1769.5	65.3	544	23	AAE22070	Francisella tulare
11	1750.5	64.6	544	21	AAY75745	Neisseria gonorrhe
12	1749	64.6	545	23	ABB99014	Detrimental organi
13	1746	64.5	545	21	AAB19080	Amino acid sequenc
14	1746	64.5	548	19	AAW65070	E. coli GroEL prot
15	1746	64.5	548	20	AAW95001	WO9902989 Seq ID 1
16	1746	64.5	548	22	AAB50536	Escherichia coli G
17	1746	64.5	548	23	ABB99010	Detrimental organi
18	1744	64.4	545	23	ABB99013	Detrimental organi
19	1743.5	64.4	544	20	AAY23905	Amino acid sequenc
20	1742	64.3	545	23	ABB99009	Detrimental organi
21	1742	64.3	545	23	ABB99012	Detrimental organi
22	1742	64.3	545	23	ABB99016	Detrimental organi
23	1741	64.3	548	19	AAW61378	GroEL N-terminal r
24	1741	64.3	548	21	AAY99621	Escherichia coli g
25	1739	64.2	545	23	ABB99015	Detrimental organi
26	1738	64.2	545	23	ABB99011	Detrimental organi
27	1736.5	64.1	548	20	AAY23908	Amino acid sequenc
28	1735	64.1	548	16	AAR67382	E. coli GroEL gene
29	1734	64.0	545	23	ABB99017	Detrimental organi
30	1733	64.0	545	20	AAY23901	Streptococcus pneu
31	1732	64.0	542	23	ABB49241	Listeria monocytog
32	1729.5	63.9	544	20	AAY23903	Streptococcus pyog
33	1723.5	63.6	546	14	AAR41661	H. pylori hsp60.
34	1723.5	63.6	546	14	AAR41200	Hsp. Helicobacter
35	1723	63.6	544	12	AAR13337	HypB protein. Chl
36	1720.5	63.5	546	19	AAW98776	H. pylori GHPO 118
37	1720	63.5	548	23	ABB99019	Detrimental organi
38	1720	63.5	548	23	ABB99021	Detrimental organi
39	1719	63.5	548	23	ABB99018	Detrimental organi
40	1718	63.4	544	16	AAR67383	C. psittaci HypB g
41	1717	63.4	548	23	ABB99020	Detrimental organi
42	1716.5	63.4	547	23	AAU76192	Human P1 protein (
43	1716.5	63.4	547	24	ABG74586	E. coli groEL prot
44	1710	63.1	544	22	AAE11757	Chlamydophila pneu
45	1710	63.1	544	23	ABB94272	Chlamydia pneumoni

ALIGNMENTS

RESULT 1
AAW16678
ID AAW16678 standard; Protein; 548 AA.
XX
AC AAW16678;
XX
DT 20-AUG-1997 (first entry)
XX
DE Lawsonia intracellularis GroEL.
XX
KW Intestinal disease; porcine proliferative enteropathy; vaccine;
KW GroEL; immunotherapy; antibody; diagnosis; heat shock protein.
XX
OS Lawsonia intracellularis.
XX
PN WO9720050-A1.
XX
PD 05-JUN-1997.
XX
PF 29-NOV-1996; 96WO-AU00767.
XX
PR 30-NOV-1995; 95AU-0006911.
PR 30-NOV-1995; 95AU-0006910.
XX
PA (DARA-) DARATECH PTY LTD.
PA (PIGR-) PIG RES & DEV CORP.
XX
PI Hasse D, Panaccio M;
XX
DR WPI; 1997-310605/28.
DR N-PSDB; AAT69201.
XX

PT Vaccine for treating or preventing Lawsonia intracellularis
PT infection - especially in pigs, containing non-pathogenic form of
PT bacterium or its components
XX
PS Claim 10; Page 38-42; 94pp; English.
XX
CC The GroEL heat shock protein (AAW16678) of Lawsonia intracellularis
CC can be used in vaccines to protect birds and animals against
CC intestinal diseases, esp. to protect pigs against porcine
CC proliferative enteropathy. It is the expression product of a DNA
CC molecule (AAT69201) obtd. by screening an L. intracellularis library
CC with rabbit anti-L. intracellularis sera. GroEL, GroES (AAW16679)
CC and other L. intracellularis polypeptides (AAW16680-85) can be
CC administered as recombinant polypeptides or expressed as
CC recombinant vaccines utilising bacterial, fungal or viral vectors.
CC Antibodies raised against the polypeptides may be useful in
CC immunotherapy, diagnosis of infection and detection.
XX
SQ Sequence 548 AA;

Query Match 100.0%; Score 2708; DB 18; Length 548;
Best Local Similarity 100.0%; Pred. No. 3.4e-170;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540

Qy        541 MGGMDGMY 548
              ||||||||
Db        541 MGGMDGMY 548
```

RESULT 2
AAY23930
ID AAY23930 standard; Protein; 545 AA.
XX
AC AAY23930;
XX
DT 22-SEP-1999 (first entry)
XX
DE Consensus mino acid sequence of a heat shock protein.
XX
KW Heat shock protein; Hsp; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS Synthetic.
XX
PN WO9935270-A1.
XX
PD 15-JUL-1999.
XX
PF 29-DEC-1998; 98WO-CA01203.
XX
PR 31-DEC-1997; 97US-0001737.
XX
PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI Mizzen L, Wisniewski J;
XX
DR WPI; 1999-430397/36.
XX
PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
PS Disclosure; Fig 10A-E; 176pp; English.
XX
CC AAY23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.
XX
SQ Sequence 545 AA;

Query Match 68.9%; Score 1864.5; DB 20; Length 545;
Best Local Similarity 69.6%; Pred. No. 1.1e-114;
Matches 380; Conservative 74; Mismatches 85; Indels 7; Gaps 5;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|:| |  :||  : |||: ||:|||||||||||||:|||||:| ||||||:|||||||
Db          2 AKDIKFGEEARRAMLRGVNALADAVKVTLGPKGRNVVLEKSFGAPTITKDGVTVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              ||||||||||:||||| ||:|:|||||||||||||||| :||:| |||| ||| ::|||||
Db         62 EDKFENMGAQLVKEVASKTNDVAGDGTTTATVLAQAIVKEGLKNVAAGANPMDLRRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              || || :||  | ||   ::||||| ||||| |  || :||||| |||| |||||||  |
Db        122 AVDAVVEELKAIAKPVETKEEIAQVATISANGDEEIGELIAEAMEKVGKEGVITVEEGKT 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||| |:|||||:|||||:||||:|: ||   ||::| ||  :|||::::|:||:||:|:
Db        182 LETELEVVEGMQFDRGYISPYFITDSEKQKAELEDPLILLTDKKISNIQDLLPVLEEVAQ 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :|||||||||||||||||||||| |:||||||||:||||||:|||||||: |
Db        242 AGKPLLIIAEDVEGEALATLVVNKLRGTLKVVAVKAPGFGDRRKAMLQDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |: ||: :|  || ||:||: |::||||||| :  |  || |||:||||::||||:
Db        302 EELGLSLEDATLEDLGQAKKVVVTKDDTTIVDGAGDAA-IAGRVAQIRSQIEESTSDYDK 360
```

```
Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
Db        361 EKLQERLAKLAGGVAVIKVGAATEVELKERKDRVEDALNATRAAVEEGIVPGGGVALLRA 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
Db        421 APALDKLKTENGDEATGVNIVLRALEAPLRQIAENAGLEGSVVVEKVKNSEAG-GYNAAT 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP--MPG--GGM 538
Db        480 GEYVDMIAAGIIDPTKVTRSALQNAASVASLMLTTEAVVVDKPE-KEAAPAGMPGMMGGM 538

Qy        539 GGMGGM 544
Db        539 GGMGGM 544
```

RESULT 3
AAY23909
ID AAY23909 standard; Protein; 549 AA.
XX
AC AAY23909;
XX
DT 22-SEP-1999 (first entry)
XX
DE Amino acid sequence of a heat shock protein.
XX
KW Heat shock protein; Hsp; immune response; immunological carrier;
KW cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS Legionella pneumophila.
XX
PN WO9935270-A1.
XX
PD 15-JUL-1999.
XX
PF 29-DEC-1998; 98WO-CA01203.
XX
PR 31-DEC-1997; 97US-0001737.
XX
PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI Mizzen L, Wisniewski J;
XX
DR WPI; 1999-430397/36.
XX
PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT useful in vaccines, as carriers for other immunogens, as anticancer
PT agents and for diagnosis
XX
PS Disclosure; Fig 10A-E; 176pp; English.
XX
CC AAY23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.
XX
SQ Sequence 549 AA;

Query Match 68.2%; Score 1848; DB 20; Length 549;
Best Local Similarity 67.3%; Pred. No. 1.4e-113;
Matches 367; Conservative 80; Mismatches 96; Indels 2; Gaps 1;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
Db          2 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
Db         62 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
Db        122 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
Db        182 LENELSVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
Db        242 SGRPLLIIAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
Db        302 EEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARITQIRAQMEETTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
Db        362 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
Db        422 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
Db        482 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 539

Qy        543 GMDGM 547
Db        540 GMGGM 544
```

RESULT 4
AAB69060
ID AAB69060 standard; Protein; 547 AA.
XX
AC AAB69060;
XX
DT 18-APR-2001 (first entry)
XX
DE Pseudomonas aeruginosa groEL protein sequence.
XX
KW Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW diagnosis; detection; infection; immune response.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200102577-A1.
XX
PD 11-JAN-2001.
XX
PF 03-JUL-2000; 2000WO-GB02554.
XX
PR 01-JUL-1999; 99GB-0015419.
XX
PA (PROV-) PROVALIS UK LTD.
XX
PI Smith CJ, Thompson SE, Smith MW, Peek K, Sizer PJH, Wilkinson MC;
XX
DR WPI; 2001-080988/09.
DR N-PSDB; AAF32452.
XX
PT Antigenic Pseudomonas aeruginosa proteins, useful in the detection

PT and/or diagnosis of P. aeruginosa infections and for producing vaccines
PT against P. aeruginosa -
XX
PS Claim 3; Fig 25; 129pp; English.
XX
CC The present invention describes antigenic Pseudomonas aeruginosa
CC proteins (P1). The P. aeruginosa proteins have antibacterial activity
CC and can be used in vaccines and as antagonists. The proteins or their
CC fragments, or antibodies are useful in the detection and/or diagnosis
CC of P. aeruginosa. They are also useful for producing a vaccine and
CC inducing an immune response against P. aeruginosa infection. An agent
CC capable of antagonising, inhibiting or otherwise interfering with the
CC function or expression of P1 are useful in the manufacture of a
CC medicament for the treatment or prophylaxis of P. aeruginosa infections.
CC The present sequence represents a specifically claimed P. aeruginosa
CC groEL protein sequence from the present invention.
XX
SQ Sequence 547 AA;

Query Match 67.4%; Score 1825.5; DB 22; Length 547;
Best Local Similarity 66.8%; Pred. No. 4.3e-112;
Matches 367; Conservative 75; Mismatches 102; Indels 5; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||:|:  ||: ||:||| ||||||||||::||||:| ||||||||||||
Db          1 MAAKEVKFGDSARKKMLVGVNVLADAVKATLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||:||:|| | :| ||||||||||||||  ||:| |||| ||| :||||
Db         61 ELKDKFENMGAQLVKDVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||| ||:  :| :: ||  | | |||||||||||| :|| |||||| |||| |||||||
Db        121 DKATVAIVAQLKELAKPCADTKAIAQVGTISANSDESIGQIIAEAMEKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
               |||  | |||||:||||||||||||  |: |  |||:| :|  :|||:::::|||:|| |
Db        181 SGLENELSVVEGMQFDRGYLSPYFVNKPDTMAAELDSPLLLLVDKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||:|||||||||||||| :|| ::| |||||||:|||||:|||||||||||||
Db        241 AKAGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: || :|  || |||||:||||||:||||   ||:||| ||| |||||:|||
Db        301 ISEEVGLSLEGATLEHLGNAKRVVINKENTTIIDGAGVQADIEARVLQIRKQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| |||||| |||||| ||||||:|||||||||:|||||| | |
Db        361 DREKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |::: :: :|  ::::  |: ::||::| ||||| |||| | |:||:||::    :||||
Db        421 RALQAIEGLKGDNEEQNVGIALLRRAVESPLRQIVANAGDEPSVVVDKVKQGSGNYGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGG--GM 538
              |:| | |:|: |::|| |||| ||| |||:  |::|||  :||  |  | |  ||  |
Db        481 ATGVYGDMIEMGILDPAKVTRSALQAAASIGGLMITTEAMVAEIVE---DKPAMGGMPDM 537

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        538 GGMGGMGGM 546
```

RESULT 5
AAR67381
ID AAR67381 standard; Protein; 547 AA.
XX
AC AAR67381;
XX
DT 25-MAR-2003 (updated)
DT 22-JUN-1995 (first entry)
XX
DE L. pneumophila HtpB gene product.
XX
KW Urease; immunogen; vaccine; diagnostic; heat shock protein; HSP;
KW GroEL-like protein; Helicobacter felis.
XX
OS Legionella pneumophila.
XX
PN WO9426901-A1.
XX
PD 24-NOV-1994.
XX
PF 19-MAY-1994; 94WO-EP01625.
XX
PR 19-MAY-1993; 93EP-0401309.
PR 19-NOV-1993; 93WO-EP03259.
XX
PA (INRM) INST NAT SANTE & RECH MEDICALE.
PA (INSP) INST PASTEUR.
XX
PI Ferrero R, Labigne A, Suerbaum S, Thiberge J;
XX
DR WPI; 1995-006797/01.
XX
PT DNA from Helicobacter pylori and Helicobacter felis - used to
PT develop prods. for detection, treatment and prevention of
PT Helicobacter infection
XX
PS Disclosure; Fig. 7A(i-vii); 168pp; English.
XX
CC The sequence of the Helicobacter pylori heat shock protein A
CC (given in AAR67374) was compared to that of other GroEL-like
CC proteins from Legionella pneumophila (AAR67381), Escherichia coli
CC (AAR67382), Chlamydia psittaci (AAR67383), Mycobacterium leprae
CC (AAR67384) and human mitochondrial protein P1 (AAR67385), and regions
CC of homology were identified.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 547 AA;

Query Match 67.2%; Score 1820.5; DB 16; Length 547;
Best Local Similarity 66.6%; Pred. No. 9.1e-112;
Matches 363; Conservative 82; Mismatches 97; Indels 3; Gaps 2;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   || ::  ||: ||:||:||:|:|||||:|||:|:| :|||||||||||
Db          2 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              | :| |||||||||||| |||| ||||||||||||::|  || | |||| ||| :||||||
Db         62 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||:||||:|  ::|| :| | |||||||||||||  || |||||| |||| ||||||:  |
Db        122 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:||||: | : | |||::|:||  :|||:|:::|| :|| |||
Db        182 LENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              |||||||||:|||||||||||| :|| ::| ||||||||:||||||:|||||| |: |
Db        242 SGRPLLIIAEDIEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  :|  ||:|||:::  | |||:|| ||: :| ||: |||||:|||:|||||
Db        302 EEIGKSLEGATLEDLGSAKRIIVTDE-TTIIDGEGKATEINARIAQIRAQMEETTSDYDR 360

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||:||| |||||| |||||| |||||| ||||||:|||||||||||| ||| | :|:
Db        361 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | || :|  :||:  |:||:||::| |:|||  ||||| |:|| || | || :|||||:
```

```
Db        421 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 480

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||| |::: |::|| ||||:|||||||||||:||||| :|:   |||:  :  | |||||
Db        481 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 538

Qy        543 GMDGM 547
              || ||
Db        539 GMGGM 543


RESULT 6
AAB69061
ID   AAB69061 standard; Protein; 574 AA.
XX
AC   AAB69061;
XX
DT   18-APR-2001  (first entry)
XX
DE   pTrcHisB expression vector groEL fusion protein sequence.
XX
KW   Pseudomonas aeruginosa; chitinase; groEL; chiA; antigen; vaccine;
KW   diagnosis; detection; infection; immune response.
XX
OS   Pseudomonas aeruginosa.
OS   Synthetic.
XX
PN   WO200102577-A1.
XX
PD   11-JAN-2001.
XX
PF   03-JUL-2000; 2000WO-GB02554.
XX
PR   01-JUL-1999;   99GB-0015419.
XX
PA   (PROV-) PROVALIS UK LTD.
XX
PI   Smith CJ,  Thompson SE,  Smith MW,  Peek K,  Sizer PJH,  Wilkinson MC;
XX
DR   WPI; 2001-080988/09.
DR   N-PSDB; AAF32453.
XX
PT   Antigenic Pseudomonas aeruginosa proteins, useful in the detection
PT   and/or diagnosis of P. aeruginosa infections and for producing vaccines
PT   against P. aeruginosa -
XX
PS   Claim 3; Fig 26; 129pp; English.
XX
CC   The present invention describes antigenic Pseudomonas aeruginosa
CC   proteins (P1). The P. aeruginosa proteins have antibacterial activity
CC   and can be used in vaccines and as antagonists. The proteins or their
CC   fragments, or antibodies are useful in the detection and/or diagnosis
CC   of P. aeruginosa. They are also useful for producing a vaccine and
CC   inducing an immune response against P. aeruginosa infection. An agent
CC   capable of antagonising, inhibiting or otherwise interfering with the
CC   function or expression of P1 are useful in the manufacture of a
CC   medicament for the treatment or prophylaxis of P. aeruginosa infections.
CC   The present sequence represents a specifically claimed pTrcHisB
CC   expression vector groEL fusion protein sequence from the present
CC   invention.
XX
SQ   Sequence   574 AA;

  Query Match           67.2%;  Score 1819;  DB 22;  Length 574;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-111;
  Matches  365;  Conservative   75;  Mismatches  103;  Indels    4;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||:|:  ||: ||:||| ||||||||||::||||:| ||||||||||||
Db         31 MAAKEVKFGDSARKKMLVGVNVLADAVKATLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 90

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||:||:|| | :| ||||||||||||||||  ||:| |||| ||| :||||
Db         91 ELKDKFENMGAQLVKDVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 150

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||| ||:  :| :: ||  | | |||||||||||||| :|| |||||| |||| |||||||
Db        151 DKATVAIVAQLKELAKPCADTKAIAQVGTISANSDESIGQIIAEAMEKVGKEGVITVEEG 210

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  | |||||:|||||||||||  |: |  |||:| :|  :|||:::::|||:|| |
Db        211 SGLENELSVVEGMQFDRGYLSPYFVNKPDTMAAELDSPLLLLVDKKISNIREMLPVLEAV 270

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||:|||||||||||||||| :|| ::| ||||||||:||||||:||||||||
Db        271 AKAGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDIAILTGGTV 330

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|  || ||||||:||||||:||||   ||:||| ||| |||||:|||
Db        331 ISEEVGLSLEGATLEHLGNAKRVVINKENTTIIDGAGVQADIEARVLQIRKQIEETTSDY 390

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| |||||| |||||| ||||||:|||||||||:||||| | |
Db        391 DREKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 450

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |::: :: :|  ::::  |: ::||::| ||||| |||| | |:||:||::    :||||
Db        451 RALQAIEGLKGDNEEQNVGIALLRRAVESPLRQIVANAGDEPSVVVDKVKQGSGNYGFNA 510

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:| | |:|: |::|| |||| ||| |||:  |::|||  :|| | | |    |||
Db        511 ATGVYGDMIEMGILDPAKVTRSALQAAASIGGLMITTEAMVAEIVEDKPAM----GGMPD 566

Qy        541 MGGMDGM 547
              |||| ||
Db        567 MGGMGGM 573


RESULT 7
AAY23915
ID   AAY23915 standard; Protein; 545 AA.
XX
AC   AAY23915;
XX
DT   22-SEP-1999  (first entry)
XX
DE   Amino acid sequence of a heat shock protein.
XX
KW   Heat shock protein; Hsp; immune response; immunological carrier;
KW   cancer control; tumour; sarcoma; cancer; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9935270-A1.
XX
PD   15-JUL-1999.
XX
PF   29-DEC-1998;   98WO-CA01203.
XX
PR   31-DEC-1997;   97US-0001737.
XX
PA   (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI   Mizzen L,  Wisniewski J;
XX
DR   WPI; 1999-430397/36.
XX
PT   New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT   useful in vaccines, as carriers for other immunogens, as anticancer
PT   agents and for diagnosis
XX
PS   Disclosure; Fig 10A-E; 176pp; English.
XX
CC   AAY23905-30 represent heat shock proteins (Hsps). The specification
CC   describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC   fragments, variants and fusion proteins, are used to elicit or enhance
```

```
CC   an immune response against Streptococcus, and to elicit a similar
CC   response to a target antigen fused to the protein. Unlike other
CC   immunological carriers, Hsp60 proteins are not immunosuppressive so
CC   provide an increased response to any conjugated or fused antigen. Also,
CC   where used for cancer control, they lack the side effects associated
CC   with endotoxins. They can also be used to detect specific antibodies
CC   and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC   breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC   polynucleotide is used for recombinant production of the protein, as
CC   a source of primers and probes for detecting streptococci in standard
CC   hybridization/amplification assays, and therapeutically in gene
CC   therapy vectors.
XX
SQ   Sequence   545 AA;

  Query Match             67.0%;  Score 1814.5;  DB 20;  Length 545;
  Best Local Similarity   66.7%;  Pred. No. 2.2e-111;
  Matches  363;  Conservative   77;  Mismatches  103;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |  : |:|:  ||: |||||:||||||||||::::|| | ||||||:|||||
Db          1 MAAKDVQFGNEVRQKMVNGVNILANAVRVTLGPKGRNVVVDRAFGGPHITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||||||| ||:|:|||||||||||||:|  ||:| | || ||  :||||
Db         61 ELKDKFENMGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |: :|| :| ||     |||||||:||||||  :| |||||| |||| ||||||:
Db        121 DKAVAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | ||  |||||||:||||||||: : || :  ||||::|  :|||::::|:||:||||
Db        181 KSLENELDVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFDKKISNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || :|||||||||||||||||||| :|| |: |||||||||:||||||:||||||||
Db        241 AKASRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLQDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|  || |||: | ||||||:|| | :  |:||| :|| |||  :|||
Db        301 ISEEVGLSLEKATLDDLGQAKRIEIGKENTTIIDGFGDAAQIEARVAEIRQQIETATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:||||||:||| |||||| |||||| ||||||||||||:|||||||||:| ||| | :
Db        361 DKEKLQERVAKLAGGVAVIKVGAATEVEMKEKKDRVEDALHATRAAVEEGVVAGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:   |:::   : |: ||: |: |::| ||||| |||| | |:|| || | |  :|:||
Db        421 RARAALENLHTGNADQDAGVQIVLRAVESPLRQIVANAGGEPSVVVNKVLEGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |||| |:|: ||:|| |||| |||:|||:| |:|||:| ||| || |  ||   |||||
Db        481 GSGEYGDMIEMGVLDPAKVTRSALQHAASIAGLMLTTDCMIAEIPEEKPAMP-DMGGMGG 539

Qy        541 MGGM 544
              ||||
Db        540 MGGM 543


RESULT 8
AAY75747
ID   AAY75747 standard; Protein; 544 AA.
XX
AC   AAY75747;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 982 protein sequence SEQ ID NO:2966.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54509.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1388; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   544 AA;

  Query Match             66.9%;  Score 1812.5;  DB 21;  Length 544;
  Best Local Similarity   66.5%;  Pred. No. 3e-111;
  Matches  362;  Conservative   78;  Mismatches  103;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |  : |:|:  ||: |||||:||||||||||::::|| | ||||||:|||||
Db          1 MAAKDVQFGNEVRQKMVNGVNILANAVRVTLGPKGRNVVVDRAFGGPHITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||||||| ||:|:|||||||||||||:|  ||:| | || ||  :||||
Db         61 ELKDKFENMGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |: :|| :| ||     |||||||:||||||  :| |||||| |||| ||||||:
Db        121 DKAVAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMQFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | ||  |||||||:||||||||: : || :  ||||::|  :|||::::|:||:||||
Db        181 KSLENELDVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFDKKISNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || :|||||||||||||||||||| :|| |: |||||||||:||||||:||||||||
Db        241 AKASRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLQDIAILTGGTV 300
```

```
Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEEVGLSIEKATLDDLGQAKRIEIGKENTTIIDGFGDAAQIEARVAEIRQQIETATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DKEKLQERVAKLAGGVAVIKVGAATEVEMKEKKDRVEDALHATRAAVEEGVVAGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RARAALENLHTGNADQDAGVQIVLRAVESPLRQIVANAGGEPSVVVNKVLEGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
Db        481 GSGEYGDMIEMGVLDPAKVTRSALQHAASIAGLMLTTDCMIAEIPEDKPAMP-DMGGMGG 539

Qy        541 MGGM 544
Db        540 MGGM 543
```

RESULT 9
ABP77245
ID ABP77245 standard; Protein; 544 AA.
XX
AC ABP77245;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 1020.
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX
DR WPI; 2003-058415/05.
DR N-PSDB; ABZ38215.
XX
PT New protein from Neisseria gonorrheae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrheae infection -
XX
PS Disclosure; Page 262; 815pp; English.
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 544 AA;

Query Match 66.7%; Score 1805.5; DB 24; Length 544;
Best Local Similarity 66.4%; Pred. No. 8.8e-111;
Matches 361; Conservative 78; Mismatches 104; Indels 1; Gaps 1;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          1 MAAKDVQFGNEVRQKMVNGVNILANAVRVTLGPKGRNVVVDRAFGGPHITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         61 ELKDKFENMGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        121 DKAVAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        181 KSLENELDVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFDKKISNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        241 AKASRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLQDIAILTGGVV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEEVGLSLEKATLDDLGQAKRIEIGKENTTVIDGFGDAAQIEARVAEIRQQIETATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DKEKLQERVAKLAGGVAVIKVGAATEVEMKEKKDRVEDALHATRAAVEEGVVAGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RARAALENLHTGNADQDAGVQIVLRAVESPLRQIVANAGGEPSVVVNKVLEGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
Db        481 GSGEYGDMIGMGVLDPAKVTRSALQHAASIAGLMLTTDCMIAEIPEEKPAVP-DMGGMGG 539

Qy        541 MGGM 544
Db        540 MGGM 543
```

RESULT 10
AAE22070
ID AAE22070 standard; Protein; 544 AA.
XX
AC AAE22070;
XX
DT 25-JUL-2002 (first entry)
XX
DE Francisella tularensis GroEL protein.
XX
KW Heat shock protein; hsp60; GroEL protein; therapy; infection; antibiotic.
XX
OS Francisella tularensis.
XX
PN WO200218600-A1.
XX
PD 07-MAR-2002.
XX
PF 24-AUG-2001; 2001WO-GB03835.
XX
PR 26-AUG-2000; 2000GB-0021008.
XX
PA (MINA) UK SEC FOR DEFENCE.
XX
PI Titball RW, Green M, Hartley MG, Mack KA, Sjoestedt A;
XX
DR WPI; 2002-362179/39.
XX
PT Novel polypeptide comprising a component of heat shock protein, Hsp60
PT obtainable from Francisella tularensis, useful for preventing or
PT treating infection caused by Francisella tularensis in an animal -
XX
PS Disclosure; Fig 1; 26pp; English.
XX
CC The invention relates to a polypeptide comprising at least one

```
CC   component of an heat shock protein 60 (hsp60) obtainable from
CC   Francisella tularensis, and in particular the GroEL protein or its
CC   fragment, and a nucleic acid sequence encoding the polypeptide. The
CC   polypeptide and polynucleotides of the invention are useful for
CC   preparing a medicament and for preventing or treating infection
CC   caused by F. tularensis in an animal. The polynucleotide of the
CC   invention is administered to an animal in a form in which it is
CC   expressed in situ. The present sequence is Francisella tularensis
CC   GroEL protein.
XX
SQ   Sequence   544 AA;

  Query Match            65.3%;  Score 1769.5;  DB 23;  Length 544;
  Best Local Similarity  64.9%;  Pred. No. 2.1e-108;
  Matches  353;  Conservative   77;  Mismatches  113;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|::||  :|| |:  ||: |||||||||||||||::||||:| ||||||||||||
Db          1 MAAKQVLFSDEARAKMLDGVNTLANAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||:|||| ||:|:|||||||||||||:  ||:| | || ||| :||||
Db         61 ELEDKFENMGAQIVKEVASKTADVAGDGTTTATVLAQALLTEGLKAVTAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||   : :||  ::||  | | | |||||||||| |:| :||:|||||| |||||||
Db        121 DKATARLVEELKALSKPCSDPKSIEQVGTISANSDATVGKLIADAMAKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              || |  |||||||:|||||||||| || | |  :|:|||||  :|||::::|:|||||
Db        181 KGFEDELDVVEGMQFDRGYLSPYFATNQENMTTDLENPYILIVDKKISNIRDLLPILEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :|  | |||||||||| ||||||||| :|| ::|  |||||||:|||||||||| |||
Db        241 SKSGRALLIIAEDVESEALATLVVNNMRGVVKVCTVKAPGFGDRRKAMLEDIATLTGATF 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              : ||  :|||  ::  |||| || : |:||||:||||: | |  |:  |:| | | :|||
Db        301 VSEDLSMKLEETNMEHLGTASRVQVTKDNTTIIDGAGEKEAIAKRINVIKANIAEANSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| ||| || |||||||||:|||:|||||||||| ||| | :
Db        361 DREKLQERLAKLSGGVAVIKVGAVTEAEMKEKKDRVDDALHATRAAVEEGIVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |: | || :  :||:  |: ::|:::| ||||| :||| | |:|| :|:  :  :|:||
Db        421 RAQKALDGLTGENDDQNHGIALLRKAIEAPLRQIVSNAGGESSVVVNQVKANQGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:  | |::: |::|| |||| |||:|||:| |::|||  | |  |    ||| ||||||
Db        481 ANDTYGDMVEMGILDPTKVTRSALQHAASIAGLMITTEAMIGEIKEAAPAMPM-GGGMGG 539

Qy        541 MGGM 544
              | ||
Db        540 MPGM 543


RESULT 11
AAY75745
ID   AAY75745 standard; Protein; 544 AA.
XX
AC   AAY75745;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 982 protein sequence SEQ ID NO:2962.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54507.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 1385; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   544 AA;

  Query Match            64.6%;  Score 1750.5;  DB 21;  Length 544;
  Best Local Similarity  64.9%;  Pred. No. 3.7e-107;
  Matches  353;  Conservative   78;  Mismatches  112;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              :||: : || :  :|:  ||: |  |  | || ||||||::::|| | ||||||:|||||
Db          1 IASQNLRFDNRFLQKMVNGVNILPAADWVALGAKGRNVVVDRAFGGPHITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||||||| ||:|:|||||||||||||:|  ||:| | || ||  :||||
Db         61 ELKDKFENMGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGMKYVTAGMNPTDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |: :|| :| ||    |||||||:||||||  :| |||||| |||| |||||:
Db        121 DKAVAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | ||  |||||||:||||||||||: : || :  ||||::|  :|||::::|:||:||||
Db        181 KSLENELDVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFVLLFDKKISNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || :|||||||||||||||||||| :|| |: |||||||||:|||||:|||||||
Db        241 AKASRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLQDIAILTGGVV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|  ||  ||: | :||||::|| | :  |:||| :|| |||  :|||
Db        301 ISEEVGLSLEKATLDDLGQTKRIEIGEENTTVIDGFGDAAQIEARVAEIRQQIETATSDY 360
```

```
Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DKEKLQERVAKLAGGVAVIKVGAATEVEMKEKKDRVEDALHATRAAVEEGVVAGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RARAALENLHTGNADQDAGVQIVLRAVESPLRQIVANAGGEPSVVVNKVLEGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
Db        481 GSGEYGDMIGMGVLDPAKVTRSALQHAASIAGLMLTTDCMIAEIPEEKPAVP-DMGGMGG 539

Qy        541 MGGM 544
Db        540 MGGM 543
```

RESULT 12
ABB99014
ID ABB99014 standard; protein; 545 AA.
XX
AC ABB99014;
XX
DT 12-DEC-2002 (first entry)
XX
DE Detrimental organism controller protein #6.
XX
KW Detrimental organism; enterobacter.
XX
OS Enterobacter aerogenes.
XX
PN JP2002223764-A.
XX
PD 13-AUG-2002.
XX
PF 31-JAN-2001; 2001JP-0023255.
XX
PR 31-JAN-2001; 2001JP-0023255.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 2002-694867/75.
DR N-PSDB; ABQ81658.
XX
PT A novel detrimental organism controller for limiting detrimental
PT organisms -
XX
PS Claim 8; Page 41-43; 78pp; Japanese.
XX
CC The invention relates to a protein that acts as a detrimental organism
CC controller and is isolated from Enterobacter aerogenes. The invention
CC encompasses a detrimental organism-resistant plant producing the protein
CC of the invention in an effective amount for controlling detrimental
CC organisms. A composition of the invention is useful for the preparation
CC of a detrimental organism controller. The current sequence represents a
CC detrimental organism controller protein of the invention.
XX
SQ Sequence 545 AA;

Query Match 64.6%; Score 1749; DB 23; Length 545;
Best Local Similarity 64.8%; Pred. No. 4.6e-107;
Matches 354; Conservative 73; Mismatches 115; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        121 DKAVVAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGDEAAIQGRVTQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RVASKIADLKGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKAGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKSDAPDLGAAGGM 538

Qy        539 GGMGGM 544
Db        539 GGMGGM 544
```

RESULT 13
AAB19080
ID AAB19080 standard; protein; 545 AA.
XX
AC AAB19080;
XX
DT 08-FEB-2001 (first entry)
XX
DE Amino acid sequence of a 60 kDa protein from Campylobacter jejuni.
XX
KW Antigenic protein; flagellaless mutant; vaccine; poultry; Campylobacter.
XX
OS Campylobacter jejuni.
XX
PN EP1043029-A1.
XX
PD 11-OCT-2000.
XX
PF 03-APR-2000; 2000EP-0201203.
XX
PR 09-APR-1999; 99EP-0201086.
XX
PA (ALKU) AKZO NOBEL NV.
XX
PI Jacobs AAC, Van Den Bosch JF, Nuijten PJM;
XX
DR WPI; 2000-589147/56.
XX
PT Novel Campylobacter antigenic proteins, useful for the production of
PT vaccines for protecting poultry form Campylobacter infection -
XX
PS Disclosure; Page 12-14; 21pp; English.
XX
CC The present sequence represents a 60 kDa protein of Campylobacter
CC jejuni. The protein is antigenic, and is visible on a Western blot
CC of C. jejuni protein after incubation of the blot with antibodies
CC against a flagellaless mutant of C. jejuni but not visible after
CC incubation with antibodies against wild type C. jejuni. The proteins
CC are used to produce vaccines which are useful for protecting animals,
CC especially poultry, against Campylobacter, especially C. jejuni.
XX

SQ Sequence 545 AA;

Query Match 64.5%; Score 1746; DB 21; Length 545;
Best Local Similarity 64.5%; Pred. No. 7.3e-107;
Matches 353; Conservative 80; Mismatches 108; Indels 6; Gaps 4;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|||:|  :|| ||  || || :|||||:||:||||:|:||||:| ||||||||||:||
Db          2 AKEIIFSDEARNKLYEGVKKLNDAVKVTMGPRGRNVLIQKSFGAPSITKDGVSVAKEVEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              :|  ||||| :|:||| ||:| |||||||||||| ||::||:: : || ||: :|||:||
Db         62 KDSLENMGASLVREVASKTADQAGDGTTTATVLAHAIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              |  |:  ||  :::  :|:|||||| ||||||| |||:||:|| |||| ||||||| |
Db        122 ACEAIVAELKKLSREVKDKKEIAQVATISANSDEKIGNLIADAMEKVGKDGVITVEEPKS 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :  |:|||||:|||||||||:|| |||  || :||||  :||||::||:||:|||: |
Db        182 INDELNVVEGMQFDRGYLSPYFITNAEKMTVELSSPYILLFDKKITNLKDLLPVLEQIQK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :|||||||:|||||||||||||||| | : ||||||||:|||||||||||||||| |
Db        242 TGKPLLIIAEDIEGEALATLVVNKLRGVLNISAVKAPGFGDRRKAMLEDIAILTGGEVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||: ::  || |  |:|||:|||||:|||:  :| ||| ||:||| ||:||||
Db        302 EELGRTLESATIQDLGQASSVIIDKDNTTIVNGAGEKANIDARVNQIKAQIAETTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| |||||| ||| ||||||||||:|||:||:|||||||| ||| | :::
Db        362 EKLQERLAKLSGGVAVIKVGATTETEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIKA 421

Qy        423 -IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFNA 480
               |:  |::    ||  |  |: |:|  |||||| |||::  :||  |   ||   ||:|
Db        422 KAKIKLDLQ---GDEAIGAAIVERALRAPLRQIAENAGFDAGVVVNSVENAKDENTGFDA 478

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              | ||| :::::|:||| || |:|| || ||||:||||  |:|  | |  ||    ||||
Db        479 AKGEYVNMLESGIIDPVKVERVALLNAVSVASMLLTTEATISEIKEDKPTMP-DMSGMGG 537

Qy        541 MGGMDGM 547
              |||| ||
Db        538 MGGMGGM 544
```

RESULT 14
AAW65070
ID AAW65070 standard; Protein; 548 AA.
XX
AC AAW65070;
XX
DT 11-SEP-1998 (first entry)
XX
DE E. coli GroEL protein.
XX
KW Chaperone; GroEL; apical region; disease; protein structure;
KW Alzheimer's disease; reconditioning; purification; inactivation;
KW denaturation; therapeutic.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Protein 1..548
FT /label= GroEL
XX
PN WO9813496-A1.
XX
PD 02-APR-1998.
XX
PF 26-SEP-1997; 97WO-GB02652.
XX
PR 03-DEC-1996; 96WO-GB02980.
PR 26-SEP-1996; 96GB-0020243.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Altamirano MM, Fersht AR, Zahn R;
XX
DR WPI; 1998-230706/20.
XX
PT New chaperone polypeptide(s) derived from GroEL - useful for, e.g.
PT refolding of proteins and for treating diseases associated with
PT protein structure such as Alzheimer's disease
XX
PS Claim 1; Fig 7; 94pp; English.
XX
CC This sequence represents the E. coli chaperone protein GroEL which can be
CC used for treating diseases associated with protein/polypeptide structure,
CC e.g. Alzheimer's disease. The sequence can also be used for
CC reconditioning molecules, especially a protein that has previously been
CC inactivated or denatured, or more generally for altering the structure of
CC a molecule by unfolding or refolding. The sequence also has applications
CC for purification procedures and can increase yield, specific activity or
CC quality of biological molecules, especially by co-expression with the
CC molecules. Therapeutic agents are administered orally, by injection or
CC intranasally, or from a slow-release formulation.
XX
SQ Sequence 548 AA;

Query Match 64.5%; Score 1746; DB 19; Length 548;
Best Local Similarity 64.5%; Pred. No. 7.3e-107;
Matches 354; Conservative 74; Mismatches 117; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |  || |: ||| ||:|||||||||||||::||||:| |||||||||:||
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| | :| |||||||||||||  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  ::  |  | |||||||||||| |:| :||||| |||| ||||||:
Db        121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||:  |||||||:||||||||:  ||    ||::|:||  :|||:::::|||:|| |
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||||||| :|| ::| ||||||||:|||||:||| ||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  || ||||||:|: |||:|| |:   |: || ||| |||| :|||
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||:|||||||||:| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    | |::  ::|:  |: :  |::| ||||  | | | |:|   |:     :|:||
Db        421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
              |: || ::|  |::|| |||| ||| ||||| |::|||| : :   || |     | |||
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        539 GGMGGMGGM 547
```

RESULT 15
AAW95001
ID AAW95001 standard; Protein; 548 AA.
XX
AC AAW95001;
XX
DT 12-MAY-1999 (first entry)
XX
DE WO9902989 Seq ID 10.
XX
KW Chaperone; molecular chaperone; groEL gene.
XX
OS Escherichia coli.
XX
PN WO9902989-A1.
XX
PD 21-JAN-1999.
XX
PF 10-JUL-1998; 98WO-GB02042.
XX
PR 28-AUG-1997; 97GB-0018261.
PR 10-JUL-1997; 97GB-0014582.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Chatellier J, Fersht A;
XX
DR WPI; 1999-121113/10.
XX
PT In vivo assay for the detection of chaperone fragment activity -
PT useful for, e.g. complementing a molecular chaperone defect in vivo
XX
PS Disclosure; Page 52-55; 64pp; English.
XX
CC The invention provides a method of providing chaperone activity in vivo,
CC by administering to a cell a fragment of a molecular chaperone which has
CC in vivo activity. Also provided is a method of determining whether a
CC fragment of a molecular chaperone is active in vivo by: (i) providing a
CC cell with a deficient molecular chaperone activity; (ii) administering
CC the fragment to the cell; and (iii) determining whether the fragment
CC complements the deficient endogenous molecular chaperone. The method can
CC determine which fragments of the molecular chaperone can be used to
CC complement a mutant or deficient molecular chaperone in vivo. The method
CC allows an in vivo assessment of chaperone activity as compared to
CC previous tests which only measured activity in vitro.
XX
SQ Sequence 548 AA;

Query Match 64.5%; Score 1746; DB 20; Length 548;
Best Local Similarity 64.5%; Pred. No. 7.3e-107;
Matches 354; Conservative 74; Mismatches 117; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   || |: |||: ||:||||||||||||||::||||:| ||||||||||:||
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| | :| ||||||||||||||||  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  :: |  | | ||||||||||||| |:| :||||| |||| ||||||:
Db        121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
               ||:  |||||||:||||||||||:  ||    ||::|:||  :|||:::::|||:|| |
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :||||||||||||||||||||| :|| ::| ||||||||:||||||:||| ||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  || ||||||:|: |||:|| |:   |: || ||| |||| :|||
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| |||||:|||||||||:| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |   | |::  ::|:  |: :  |::| ||||| | | | |:|   |:     :|:||
Db        421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
              |: || ::|  |::|| |||| ||| ||||| |::|||| : :   || |    | |||
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        539 GGMGGMGGM 547
```

Search completed: January 28, 2004, 13:15:54
Job time : 49 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:15:00 ; Search time 22 Seconds
(without alignments)
1053.925 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1844	68.1	548	2	US-08-467-822-31	Sequence 31, Appl
2	1844	68.1	548	3	US-08-432-697-31	Sequence 31, Appl
3	1844	68.1	548	3	US-08-466-248-31	Sequence 31, Appl
4	1820.5	67.2	551	4	US-09-252-991A-24209	Sequence 24209, A
5	1768.5	65.3	550	4	US-09-328-352-5508	Sequence 5508, Ap
6	1741	64.3	548	2	US-08-467-822-32	Sequence 32, Appl
7	1741	64.3	548	3	US-09-472-971-3	Sequence 3, Appli
8	1741	64.3	548	3	US-08-432-697-32	Sequence 32, Appl
9	1741	64.3	548	3	US-08-466-248-32	Sequence 32, Appl
10	1723.5	63.6	546	3	US-08-470-260-6	Sequence 6, Appli
11	1723.5	63.6	546	3	US-08-471-491-6	Sequence 6, Appli
12	1723.5	63.6	546	3	US-08-466-662-6	Sequence 6, Appli
13	1723.5	63.6	546	4	US-08-256-847C-1	Sequence 1, Appli
14	1723.5	63.6	546	4	US-08-256-847C-7	Sequence 7, Appli
15	1721	63.6	544	2	US-08-467-822-33	Sequence 33, Appl
16	1721	63.6	544	3	US-08-432-697-33	Sequence 33, Appl
17	1721	63.6	544	3	US-08-466-248-33	Sequence 33, Appl
18	1716.5	63.4	547	4	US-08-461-722-2	Sequence 2, Appli
19	1716.5	63.4	547	4	US-08-336-251-2	Sequence 2, Appli
20	1716.5	63.4	547	4	US-09-468-041-2	Sequence 2, Appli
21	1716.5	63.4	547	5	PCT-US94-06362-2	Sequence 2, Appli
22	1708	63.1	544	4	US-09-198-452A-153	Sequence 153, App
23	1702	62.9	545	2	US-08-467-822-30	Sequence 30, Appl
24	1702	62.9	545	3	US-08-432-697-30	Sequence 30, Appl
25	1702	62.9	545	3	US-08-466-248-30	Sequence 30, Appl
26	1683.5	62.2	641	4	US-09-613-303-51	Sequence 51, Appl
27	1597	59.0	541	2	US-08-997-080-160	Sequence 160, App
28	1597	59.0	541	2	US-08-997-362-160	Sequence 160, App
29	1597	59.0	541	3	US-09-095-855-160	Sequence 160, App
30	1597	59.0	541	4	US-09-324-542-160	Sequence 160, App
31	1597	59.0	541	4	US-09-205-426-160	Sequence 160, App
32	1595.5	58.9	540	4	US-09-613-303-4	Sequence 4, Appli
33	1595.5	58.9	639	4	US-09-613-303-17	Sequence 17, Appl
34	1595.5	58.9	648	4	US-09-613-303-29	Sequence 29, Appl
35	1595.5	58.9	948	4	US-09-613-303-21	Sequence 21, Appl
36	1591.5	58.8	540	4	US-08-461-722-4	Sequence 4, Appli
37	1591.5	58.8	540	4	US-08-336-251-4	Sequence 4, Appli
38	1591.5	58.8	540	4	US-09-468-041-4	Sequence 4, Appli
39	1591.5	58.8	540	5	PCT-US94-06362-4	Sequence 4, Appli
40	1582.5	58.4	540	4	US-09-134-001C-4705	Sequence 4705, Ap
41	1577.5	58.3	541	2	US-08-467-822-34	Sequence 34, Appl
42	1577.5	58.3	541	2	US-08-447-154-19	Sequence 19, Appl
43	1577.5	58.3	541	3	US-08-432-697-34	Sequence 34, Appl
44	1577.5	58.3	541	3	US-08-466-248-34	Sequence 34, Appl
45	1575	58.2	540	2	US-08-368-834-20	Sequence 20, Appl

ALIGNMENTS

```
RESULT 1
US-08-467-822-31
; Sequence 31, Application US/08467822
; Patent No. 5843460
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,822
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-822-31

  Query Match             68.1%;  Score 1844;  DB 2;  Length 548;
  Best Local Similarity   67.2%;  Pred. No. 1e-152;
  Matches  366;  Conservative   81;  Mismatches   96;  Indels    2;  Gaps    1;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   || ::  ||: ||:||:||:||:|||||:|||:|:| :|||||||||||
Db          2 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              | :| |||||||||||| |||| |||||||||||||::|  || | |||| ||| :||||||
Db         62 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||:||||:|  ::|| :| | ||||||||||||  || ||||||  |||| ||||||:  |
Db        122 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:|||| | : | |||::|:||  :||::|:::|| :|| |||
Db        182 LENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              |||||||||:|||||||||| :|| ::| |||||||:||||||:||||||  |: |
Db        242 SGRPLLIIAEDIEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  :|  ||:|||:|: ||||||:|| ||: :| ||: |||||:|||:|||||
Db        302 EEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||:||| |||||| |||||| |||||| ||||||:|||||||||||| ||| | :|:
Db        362 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | || :|  :||:  |:||:||::| |:|||  ||||| |:|| || | || :|||||:
Db        422 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||| |::: |::|| ||||:||||||||||:||||| :|:   |||:  :  | |||||
Db        482 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 539

Qy        543 GMDGM 547
              || ||
Db        540 GMGGM 544


RESULT 2
US-08-432-697-31
; Sequence 31, Application US/08432697
; Patent No. 6248330
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-432-697-31

  Query Match             68.1%;  Score 1844;  DB 3;  Length 548;
  Best Local Similarity   67.2%;  Pred. No. 1e-152;
  Matches  366;  Conservative   81;  Mismatches   96;  Indels    2;  Gaps    1;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   || ::  ||: ||:||:||:||:|||||:|||:|:| :|||||||||||
Db          2 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              | :| |||||||||||| |||| |||||||||||||::|  || | |||| ||| :||||||
Db         62 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||:||||:|  ::|| :| | ||||||||||||  || ||||||  |||| ||||||:  |
Db        122 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:|||| | : | |||::|:||  :||::|:::|| :|| |||
Db        182 LENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              |||||||||:|||||||||| :|| ::| |||||||:||||||:||||||  |: |
Db        242 SGRPLLIIAEDIEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  :|  ||:|||:|: ||||||:|| ||: :| ||: |||||:|||:|||||
Db        302 EEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||:||| |||||| |||||| |||||| ||||||:|||||||||||| ||| | :|:
Db        362 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | || :|  :||:  |:||:||::| |:|||  ||||| |:|| || | || :|||||:
Db        422 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||| |::: |::|| ||||:||||||||||:||||| :|:   |||:  :  | |||||
Db        482 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 539

Qy        543 GMDGM 547
              || ||
Db        540 GMGGM 544


RESULT 3
US-08-466-248-31
; Sequence 31, Application US/08466248
; Patent No. 6258359
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,248
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-248-31

  Query Match             68.1%;  Score 1844;  DB 3;  Length 548;
  Best Local Similarity   67.2%;  Pred. No. 1e-152;
  Matches  366;  Conservative   81;  Mismatches   96;  Indels    2;  Gaps    1;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   || ::  ||: ||:||:||:||:|||||:|||:|:| :||||||||||||
Db          2 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              | :| |||||||||||| |||| ||||||||||||::|  || | |||| ||| :||||||
Db         62 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||:||||:|  ::|| :| | |||||||||||  || ||||||  |||| |||||||:  |
Db        122 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:||||: | : | |||::|:||  :||::|:::|| :|| |||
Db        182 LENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              |||||||||:|||||||||||| :|| ::| ||||||||:||||||:|||||| |: |

Db        242 SGRPLLIIAEDIEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  :|  ||:|||:|: ||||||:|| ||: :| ||: |||||:|||:|||||
Db        302 EEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||:||| |||||| |||||| |||||| ||||||:|||||||||| ||| | :|:
Db        362 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | || :|  :||:  |:||:||::| |:||| ||||| |:|| || | || :|||||:
Db        422 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||| |::: |::|| ||||:|||||||||||:||||| :|:   |||:  :  | |||||
Db        482 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 539

Qy        543 GMDGM 547
              || ||
Db        540 GMGGM 544


RESULT 4
US-09-252-991A-24209
; Sequence 24209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 24209
;   LENGTH:  551
;   TYPE:  PRT
;   ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-24209

  Query Match             67.2%;  Score 1820.5;  DB 4;  Length 551;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-150;
  Matches  366;  Conservative   75;  Mismatches  103;  Indels    5;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||:|:  ||: ||:||| ||||||||||::||||:| ||||||||||||
Db          5 MAAKEVKFGDSARKKMLVGVNVLADAVKATLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 64

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||:||:|| | :| |||||||||||||||  ||:| |||| ||| :||||
Db         65 ELKDKFENMGAQLVKDVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 124

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||| ||:  :| :: ||  | | |||||||||||| :|| |||||| |||| |||||||
Db        125 DKATVAIVAQLKELAKPCADTKAIAQVGTISANSDESIGQIIAEAMEKVGKEGVITVEEG 184

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||| | |||||:|||||||||||  |: |  |||:|  :|  :|||:::::(((:|{ {
Db        185 SGLENELSVVEGMQFDRGYLSPYFVNKPDTMAAELDSPLLLLVDKKISNIREMLPVLEAV 244

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||:|||||||||||||| :|| ::| ||||||||:|||||:|||||||||||
Db        245 AKAGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDIAILTGGTV 304

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|  || ||||||:||||||:||||   ||:||| ||| |||||:|||
Db        305 ISEEVGLSLEGATLEHLGNAKRVVINKENTTIIDGAGVQADIEARVLQIRKQIEETTSDY 364
```

```
Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        365 DREKLQERLAKLAGGVAVIKVGAATEVEMKENKARVEDALHATRAAVEEGVVPGGGVALV 424

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        425 RALQAIEGLKGDNEEQNVGIALLRRAVESPLRQIVANAGDEPSVVVDKVKQGSGNYGFNA 484

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGG--GM 538
Db        485 ATGVYGDMIEMGILDPAKVTRSALQAAASIGGLMITTEAMVAEIVE---DKPAMGGMPDM 541

Qy        539 GGMGGMDGM 547
Db        542 GGMGGMGGM 550


RESULT 5
US-09-328-352-5508
; Sequence 5508, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5508
;   LENGTH: 550
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5508

  Query Match             65.3%;  Score 1768.5;  DB 4;  Length 550;
  Best Local Similarity   64.0%;  Pred. No. 4.1e-146;
  Matches  348;  Conservative   87;  Mismatches  108;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          7 MSAKDVKFGDSARSKMIAGVNVLADAVKVTLGPKGRNVVIDRSFGAPHITKDGVTVAKEI 66

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         67 SLKDKFENMGAQLVREVSSKTNDIAGDGTTTATVLAQAILNEGIKSVTAGMNPMDLKRGI 126

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        127 DIAVKTVVENIRSIAKPADDFKAIEQVGSISANSDTTVGKLIAQAMEKVGKEGVITVEEG 186

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        187 SGFEDALDVVEGMQFDRGYISPYFANKQDTLTAELENPFILLVDKKISNIRELISVLEAV 246

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        247 AKTGKPLLIIAEDVEGEALATLVVNNMRGIIKVCAVKAPGFGDRRKAMLQDIAILTGATV 306

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        307 ISEEVGMSLEQATLQDLGTAHKITVSKENTVIVDGAGDAAAIAERVQQIRAQIEESTSEY 366

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        367 DREKLQERVAKLAGGVAVIKIGAATEVEMKEKKDRVDDALHATRAAVEEGVVAGGGVALV 426

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        427 RAVNALEGLKGANEDQTAGINILRRAIEAPLRQIVANAGDEPSVVINAVKNGEGNFGYNA 486

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
Db        487 ATGEYGDMLEMGILDPAKVTRSALEHAASVAGLMLTTECMITDIPEDKPAAP-DMGGMGG 545

Qy        541 MGGM 544
Db        546 MGGM 549


RESULT 6
US-08-467-822-32
; Sequence 32, Application US/08467822
; Patent No. 5843460
;  GENERAL INFORMATION:
;    APPLICANT:  Labigne, Agnes
;    APPLICANT:  Sauerbaum, Sebastien
;    APPLICANT:  Ferrero, Richard L.
;    APPLICANT:  Thiberge, Jean-Michel
;    TITLE OF INVENTION:  IMMUNOGENIC COMPOSITIONS AGAINST
;    TITLE OF INVENTION:  HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;    TITLE OF INVENTION:  COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,822
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/447,177
;      FILING DATE:  19-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/432,697
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0137-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  548 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-822-32

  Query Match             64.3%;  Score 1741;  DB 2;  Length 548;
  Best Local Similarity   64.3%;  Pred. No. 1e-143;
  Matches  353;  Conservative   74;  Mismatches  118;  Indels    4;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120
```

```
Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        241 AKAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy        539 GGMGGMDGM 547
Db        539 GGMGGMGGM 547
```

RESULT 7
US-09-472-971-3
; Sequence 3, Application US/09472971
; Patent No. 6197547
; GENERAL INFORMATION:
; APPLICANT: SOGO, Kazuyo
; APPLICANT: YANAGI, Hideki
; APPLICANT: YURA, Takashi
; TITLE OF INVENTION: TRIGGER FACTOR EXPRESSION PLASMIDS
; FILE REFERENCE: 1422-409P
; CURRENT APPLICATION NUMBER: US/09/472,971
; CURRENT FILING DATE: 1999-12-28
; EARLIER APPLICATION NUMBER: JP10-372965
; EARLIER FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-472-971-3

Query Match 64.3%; Score 1741; DB 3; Length 548;
Best Local Similarity 64.3%; Pred. No. 1e-143;
Matches 353; Conservative 74; Mismatches 118; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        241 AKAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db        421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy        539 GGMGGMDGM 547
Db        539 GGMGGMGGM 547
```

RESULT 8
US-08-432-697-32
; Sequence 32, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-432-697-32

Query Match 64.3%; Score 1741; DB 3; Length 548;
Best Local Similarity 64.3%; Pred. No. 1e-143;
Matches 353; Conservative 74; Mismatches 118; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   || |: |||: ||:||||||||||||||::||||:| ||||||||:||
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| | :| ||||||||||||||  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  :: |  | | ||||||||||||| |:| :||||| |||| ||||||:
Db        121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||:  |||||||:||||||||:  ||    ||::|:||  :|||:::::|||:|| |
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||| ||| :|| ::| ||||||||:||||||:||| ||||
Db        241 AKAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  || ||||||:|: |||:|| |:  |: || ||| |||| :|||
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||:|||||||:| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    | |::  ::|:  |: :  |::| |||||  | | | |:|   |:     :|:||
Db        421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
              |: || ::|  |::|| |||| ||| ||||| |::|||| : :   || |   | |||
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        539 GGMGGMGGM 547
```

RESULT 9
US-08-466-248-32
; Sequence 32, Application US/08466248
; Patent No. 6258359
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,248
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-248-32

Query Match 64.3%; Score 1741; DB 3; Length 548;
Best Local Similarity 64.3%; Pred. No. 1e-143;
Matches 353; Conservative 74; Mismatches 118; Indels 4; Gaps 2;

Qy 1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db 1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy 61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db 61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI 120

Qy 121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db 121 DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy 181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db 181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy 241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db 241 AKAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy 301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db 301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY 360

Qy 361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db 361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy 421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
Db 421 RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA 480

Qy 481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
Db 481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKNDAADLGAAGGM 538

Qy 539 GGMGGMDGM 547
Db 539 GGMGGMGGM 547

```
RESULT 10
US-08-470-260-6
; Sequence 6, Application US/08470260
; Patent No. 6077706
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    APPLICANT:  Bugnoli, Massimo
;    APPLICANT:  Telford, John
;    APPLICANT:  Macchia, Giovanni
;    APPLICANT:  Rappuoli, Rino
;    TITLE OF INVENTION:  Helicobacter Pylori Proteins Useful
;    TITLE OF INVENTION:  for Vaccines and Diagnostics
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,260
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/256,848
;      FILING DATE:  21-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0316.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2708
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  546 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-470-260-6

  Query Match             63.6%;  Score 1723.5;  DB 3;  Length 546;
  Best Local Similarity   64.1%;  Pred. No. 3.5e-142;
  Matches  350;  Conservative   73;  Mismatches  120;  Indels    3;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||| |   ||  |  || :| :|||||:||:||||:|:||:|:| ||||||||||||||
Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
                 |||||:|||| ||:| |||||||||||| :|::||:: : || ||: :|||:||
Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              |  |:  ||   :|    ::|| || ||||||  || :||:|| |||| ||||||||||
Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :|  |||||||:|||||||||||||| |||  :||| |||  :|||:||||:||:||:  |
Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
                :|||||||:||||| |||||||| | : ||||||||:||| ||:||||||||: |
Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |: |||  :  || | |:||||:||||||| | |:|:| || ||: ||  |:||||:
Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||||| |||||| ||||:| |||||||||:|||:||:|||||||| ||| | :|:
Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
               : :      |||  |  || |::: || ||| ||||:| :|| :| : :  ||||::
Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541
              |:| |: | |:||| || ||||||| ||:|||||||  : |  | |    ||  ||||||
Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKATPAMPDMGGMGGM 539

Qy        542 GGMDGM 547
              ||| ||
Db        540 GGMGGM 545


RESULT 11
US-08-471-491-6
; Sequence 6, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0044
;  CURRENT APPLICATION NUMBER: US/08/471,491B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-08-471-491-6

  Query Match             63.6%;  Score 1723.5;  DB 3;  Length 546;
  Best Local Similarity   64.1%;  Pred. No. 3.5e-142;
  Matches  350;  Conservative   73;  Mismatches  120;  Indels    3;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||| |   ||  |  || :| :|||||:||:||||:|:||:|:| ||||||||||||||
Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
                 |||||:|||| ||:| |||||||||||| :|::||:: : || ||: :|||:||
Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              |  |:  ||   :|    ::|| || ||||||  || :||:|| |||| ||||||||||
Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :|  |||||||:|||||||||||||| |||  :||| |||  :|||:||||:||:||:  |
Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
                :|||||||:||||| |||||||| | : ||||||||:||| ||:||||||||: |
Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |: |||  :  || | |:||||:||||||| | |:|:| || ||: ||  |:||||:
Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
```

```
Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482

Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541

Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKATPAMPDMGGMGGM 539

Qy        542 GGMDGM 547

Db        540 GGMGGM 545


RESULT 12
US-08-466-662-6
; Sequence 6, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0057
;  CURRENT APPLICATION NUMBER: US/08/466,662B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-08-466-662-6

  Query Match           63.6%;  Score 1723.5;  DB 3;  Length 546;
  Best Local Similarity  64.1%;  Pred. No. 3.5e-142;
  Matches  350;  Conservative   73;  Mismatches  120;  Indels    3;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62

Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122

Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182

Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242

Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302

Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362

Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422

Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482

Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541

Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKATPAMPDMGGMGGM 539

Qy        542 GGMDGM 547

Db        540 GGMGGM 545


RESULT 13
US-08-256-847C-1
; Sequence 1, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
;  APPLICANT: Rappuoli, Rino
;  APPLICANT:  Costantino, Paolo
;  APPLICANT:  No. 6403099elli, Francesco
;  TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
;  FILE REFERENCE: CHIR-0042
;  CURRENT APPLICATION NUMBER: US/08/256,847C
;  CURRENT FILING DATE:  1994-11-01
;  PRIOR APPLICATION NUMBER: PCT/EP93/00516
;  PRIOR FILING DATE: 1993-03-08
;  PRIOR APPLICATION NUMBER: FI92A000058
;  PRIOR FILING DATE: 1992-03-06
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 546
;   TYPE: PRT
;   ORGANISM: H. pylori
US-08-256-847C-1

  Query Match           63.6%;  Score 1723.5;  DB 4;  Length 546;
  Best Local Similarity  64.1%;  Pred. No. 3.5e-142;
  Matches  350;  Conservative   73;  Mismatches  120;  Indels    3;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62

Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122

Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182

Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242

Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302

Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362

Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422

Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482

Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541
```

```
Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKATPAMPDMGGMGGM 539

Qy        542 GGMDGM 547
              ||| ||
Db        540 GGMGGM 545
```

RESULT 14
US-08-256-847C-7
; Sequence 7, Application US/08256847C
; Patent No. 6403099
; GENERAL INFORMATION:
; APPLICANT: Rappuoli, Rino
; APPLICANT: Costantino, Paolo
; APPLICANT: No. 6403099elli, Francesco
; TITLE OF INVENTION: Conjugates Formed From Heat Shock Proteins And Oligo or Polysacch
; FILE REFERENCE: CHIR-0042
; CURRENT APPLICATION NUMBER: US/08/256,847C
; CURRENT FILING DATE: 1994-11-01
; PRIOR APPLICATION NUMBER: PCT/EP93/00516
; PRIOR FILING DATE: 1993-03-08
; PRIOR APPLICATION NUMBER: FI92A000058
; PRIOR FILING DATE: 1992-03-06
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 546
; TYPE: PRT
; ORGANISM: H. pylori
US-08-256-847C-7

Query Match 63.6%; Score 1723.5; DB 4; Length 546;
Best Local Similarity 64.1%; Pred. No. 3.5e-142;
Matches 350; Conservative 73; Mismatches 120; Indels 3; Gaps 2;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||| |   ||  |  || :| :|||||:||:||||:|:||:|:| ||||||||||||||
Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
                |||||:||||| ||:| |||||||||||| :|::||:: : || ||: :|||:||
Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              |  |:  ||   :|    ::|| || |||||||  || :||:|| |||| ||||||||||
Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :|  |||||||:|||||||||||||| |||  :||| |||  :|||:||||:||:||:  |
Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :|||||||:|||| |||||||| | : |||||||:||| ||:|||||||: |
Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |: |||  :  || | |:||||:||||||| | |:|:| || ||: ||  |:||||:
Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| |||||| ||||:| |||||||||:|||:||:|||||||| ||| | :|:
Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              : :       |||  |  || |::: || ||| |||||:| :|| :| : :  |||||::
Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541
              |:|  |: | |:||| || |||||| ||:|||||||  : |  | |    ||  ||||||
Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKATPAMPDMGGMGGM 539

Qy        542 GGMDGM 547
              ||| ||
```

```
Db        540 GGMGGM 545
```

RESULT 15
US-08-467-822-33
; Sequence 33, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-822-33

Query Match 63.6%; Score 1721; DB 2; Length 544;
Best Local Similarity 64.8%; Pred. No. 5.7e-142;
Matches 351; Conservative 73; Mismatches 114; Indels 4; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:| | ::  ||:|: :||  || |||||||||:|||:|||||| :|||||:|||||
Db          1 MAAKNIKYNEDARKKIHKGVKTLAEAVKVTLGPKGRHVVIDKSFGSPQVTKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||| |||||||||||| ||:| |||||||||||:||| ||:: | || |||  ||||
Db         61 ELEDKHENMGAQMVKEVASKTADKAGDGTTTATVLAEAIYSEGLRNVTAGANPMLDKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||  |  |:  |:|| :  |||||| ||||:|  |||:||||| |||| | ||||||
Db        121 DKAVKVVVDEIKKISKPVQHHKEIAQVATISANNDAEIGNLIAEAMEKVGKNGSITVEEA 180
```

```
Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              || || ||||||| |:||||| || ||||    | |:   :|  :|||: :|| ||:|:||
Db        181 KGFETVLDVVEGMNFNRGYLSSYFSTNPETQECVLEEALVLIYDKKISGIKDFLPVLQQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              |:  |||||||||:||||||||||:||   :| ||||||||:|||||||||||||||:
Db        241 AESGRPLLIIAEDIEGEALATLVVNRLRAGFRVCAVKAPGFGDRRKAMLEDIAILTGGQL 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |:|||| :|: || ||:|:: ||:||||:| |  .|||::| : |: |||:::|||
Db        301 ISEELGMKLENTTLAMLGKAKKVIVSKEDTTIVEGLGSKEDIESRCESIKKQIEDSTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:|||||||||| |||||| |||||| |||||||||:|| :|| |||||||:|||||| |
Db        361 DKEKLQERLAKLSGGVAVIRVGAATEIEMKEKKDRVDDAQHATLAAVEEGILPGGGTALV 420

Qy        421 RSIKVLDDIKP--ADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGF 478
              | |  |:   |   ::||  |  |: ::|  ||:|||||| ||:|: ::|       |:
Db        421 RCIPTLEAFIPILTNEDEQIGARIVLKALSAPLKQIAANAGKEGAIICQQVLSRSSSEGY 480

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDM--PMPGG 536
              :|    | |:|:||::|| |||| ||::|||| |||||  ||: || |      |||
Db        481 DALRDAYTDMIEAGILDPTKVTRCALESAASVAGLLLTTEALIADIPEEKSSSAPAMPGA 540

Qy        537 GM 538
              ||
Db        541 GM 542
```

Search completed: January 28, 2004, 13:18:23
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:17:21 ; Search time 38 Seconds
(without alignments)
2981.360 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1862.5	68.8	547	12	US-10-369-493-17524	Sequence 17524, A
2	1857	68.6	547	12	US-10-369-493-10014	Sequence 10014, A
3	1840.5	68.0	544	12	US-10-369-493-11760	Sequence 11760, A
4	1839.5	67.9	547	12	US-10-369-493-7917	Sequence 7917, Ap
5	1835.5	67.8	542	12	US-10-369-493-10295	Sequence 10295, A
6	1834	67.7	541	12	US-10-369-493-9287	Sequence 9287, Ap
7	1833	67.7	544	12	US-10-369-493-15468	Sequence 15468, A
8	1833	67.7	544	12	US-10-369-493-16217	Sequence 16217, A
9	1828.5	67.5	543	12	US-10-369-493-15265	Sequence 15265, A
10	1826	67.4	542	12	US-10-369-493-11922	Sequence 11922, A
11	1812.5	66.9	551	12	US-10-369-493-12330	Sequence 12330, A
12	1811	66.9	540	12	US-10-369-493-15838	Sequence 15838, A
13	1810	66.8	552	12	US-10-369-493-12170	Sequence 12170, A
14	1786	66.0	545	12	US-10-369-493-167	Sequence 167, App
15	1780.5	65.7	544	12	US-10-369-493-17132	Sequence 17132, A
16	1778.5	65.7	545	12	US-10-228-167A-2	Sequence 2, Appli
17	1771.5	65.4	543	12	US-10-369-493-11916	Sequence 11916, A
18	1770	65.4	536	12	US-10-369-493-4511	Sequence 4511, Ap
19	1770	65.4	540	12	US-10-369-493-7269	Sequence 7269, Ap
20	1746	64.5	545	15	US-10-192-419-2	Sequence 2, Appli
21	1746	64.5	548	11	US-09-415-849-1	Sequence 1, Appli
22	1746	64.5	548	11	US-09-276-455-10	Sequence 10, Appl
23	1743.5	64.4	544	12	US-10-369-493-23094	Sequence 23094, A
24	1737.5	64.2	547	12	US-10-369-493-10862	Sequence 10862, A
25	1721	63.6	521	12	US-10-369-493-9597	Sequence 9597, Ap
26	1720.5	63.5	546	12	US-09-882-227-418	Sequence 418, App
27	1716.5	63.4	540	12	US-10-369-493-181	Sequence 181, App
28	1716.5	63.4	547	15	US-10-046-649-2	Sequence 2, Appli
29	1710	63.1	544	9	US-09-841-132-400	Sequence 400, App
30	1708.5	63.1	548	12	US-10-369-493-513	Sequence 513, App
31	1708	63.1	544	12	US-10-289-762-153	Sequence 153, App
32	1698.5	62.7	542	12	US-10-369-493-17907	Sequence 17907, A
33	1693	62.5	551	12	US-10-369-493-20184	Sequence 20184, A
34	1683.5	62.2	641	15	US-10-267-311-51	Sequence 51, Appl
35	1658	61.2	551	12	US-10-369-493-18994	Sequence 18994, A
36	1650	60.9	544	10	US-09-759-272B-5	Sequence 5, Appli
37	1638.5	60.5	576	12	US-10-369-493-8322	Sequence 8322, Ap
38	1636	60.4	541	15	US-10-156-761-11465	Sequence 11465, A
39	1635	60.4	538	12	US-10-369-493-2950	Sequence 2950, Ap
40	1628.5	60.1	542	15	US-10-156-761-12526	Sequence 12526, A
41	1610.5	59.5	544	12	US-10-369-493-20276	Sequence 20276, A
42	1597	59.0	541	11	US-09-880-505-160	Sequence 160, App
43	1597	59.0	541	12	US-10-205-979-40	Sequence 40, Appl
44	1597	59.0	541	14	US-10-051-643-160	Sequence 160, App
45	1595.5	58.9	540	10	US-09-847-637B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-369-493-17524
; Sequence 17524, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17524
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-17524

Query Match 68.8%; Score 1862.5; DB 12; Length 547;
Best Local Similarity 66.8%; Pred. No. 1.3e-138;
Matches 366; Conservative 80; Mismatches 97; Indels 5; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|||:|  ||| ::  ||: |||||| |||||||:||::||||||:||||||||||||
Db          1 MAAKEIIFSEKARSRMVHGVNLLANAVKATLGPKGRHVVLDKSFGSPIITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||:|||| ||:| |||||||||||||: ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMLKEVASKTNDHAGDGTTTATVLAQALIREGCKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||:|   ||  |:||| | | |||| ||||||| :||||||||| |||| ||||:||
```

```
Db        121 DKAVIAAVTELKKISKPTSDDKAIAQVATISANSDESIGNIIAEAMKKVGKEGVITIEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 TTLENELDVVEGMQFDRGYSSPYFINNQQSQIVELDNPYILLHDKKISSVRDLLTVLDAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 AKESKPLLIVAEEVEGEALATLVVNNIRGIIKVCAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEEVGLSLEKATTSHLGKAKKVRVSKENTTIIDGIGDNDAINGRVKQIKTQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVDDALLATRAAVEEGVIPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Db        421 RAITAISNLKGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGKDNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMP---GGG 537

Db        481 ATGEFGDMVNLGILDPTKVTRSALQNAASIAGLMITTEAMVAE--APKKDEPTPPAAGGG 538

Qy        538 MGGMGGMD 545

Db        539 MGGMGGMD 546
```

RESULT 2
US-10-369-493-10014
; Sequence 10014, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10014
; LENGTH: 547
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-10014

Query Match 68.6%; Score 1857; DB 12; Length 547;
Best Local Similarity 69.5%; Pred. No. 3.4e-138;
Matches 382; Conservative 72; Mismatches 90; Indels 6; Gaps 4;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKEVKFGEAARAKMLNGVNILANAVKVTLGPKGRNVVLDKSWGAPRMTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELEDKFENMGAQMVREVSSKTADVAGDGTTTATVLAQAIIREGMKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DLAVEAVVVGLKGISREVANSQEIAQVGAISANSDKVVGDMIAEAMDKVGKEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 KGLETTLDVVEGMQFDRGYLSPYFVTNADKMLVQMENPLILLVEKKISNLQQILQILEGA 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 VQSSRPLMIIAEDVEGEALATLVVNKLRGGLKVCAVKAPGFGDRRKAMMEDIATLTGGVL 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 VSEDVGVKLENVTMDMLGMAKSIVVTKEDTTIIDGAGDHEAIKARVNQIRAQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERLAKLAGGVAVIKVGGATEVEVKERKDRVDDALHATRAAVEEGIVPGGGVALL 420

Qy        421 RSIKV-LDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGF 478

Db        421 RAREASLTNLQGANHDQQVGINIVTRALEEPLRIIASNAGAEGSVVVNRVVETKETNFGF 480

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGG 537

Db        481 NAATGVYEDLVASGVIDPAKVVRHALQAAASVAGLMITTEAMVAELP---KDEPAMPGGD 537

Qy        538 MGGMGGMDGM 547

Db        538 MGGMGGMGGM 547
```

RESULT 3
US-10-369-493-11760
; Sequence 11760, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11760
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11760

Query Match 68.0%; Score 1840.5; DB 12; Length 544;
Best Local Similarity 69.2%; Pred. No. 6.9e-137;
Matches 376; Conservative 69; Mismatches 95; Indels 3; Gaps 3;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKEVKFGRSAREKMLKGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DLAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
```

```
Db        241 VQTGKPLVIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEDLGIKLESVTLDMLGKSKKVSISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479

Db        421 RSSTKI-TVKGVNDDQEAGINIVRKALQSLVRQIAENAGDEASIVVGKILDKNEDNYGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538

Db        480 AQTGEYGDLIALGIVDPVKVVRTALQNAASVASLLITTEAMIAELPKKESAMPQMPGGGM 539

Qy        539 GGM 541

Db        540 GGM 542


RESULT 4
US-10-369-493-7917
; Sequence 7917, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7917
;   LENGTH: 547
;   TYPE: PRT
;   ORGANISM: Rhodobacter sphaeroides
US-10-369-493-7917

  Query Match           67.9%;  Score 1839.5;  DB 12;  Length 547;
  Best Local Similarity  67.9%;  Pred. No. 8.3e-137;
  Matches  372;  Conservative   70;  Mismatches  103;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKDVKFDTDARDRMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELSDKFENMGAQMVKEVASRTNDEAGDGTTTATVLAQAIIKEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DLATSKVVEAIKAAARPVNDSHEVAQVGTISANGEAQIGRFIADAMQKVGNEGVITVEEN 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 KGLETEVEVVEGMQFDRGYLSPYFVTNADKMTAELDDVYILLHEKKLSSLQPMVPLLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 IQSQKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLQDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEDLGMKLENVTIDMLGRAKKISINKDNTTIVDGNGDKAEIDARVAQIRNQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERVAKLAGGVAVIRVGGMTEVEVKERKDRVDDALNATRAAVQEGIVVGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479

Db        421 QGGKALDGLTGENPDQNAGITIVRRALEAPLRQIAQNAGVDGSVVAGKVRESNEKSFGFN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539

Db        481 AQTEEYGDMFKFGVIDPAKVVRTALEDAASVASLLITTEAMIADKPEPKS--PAGGPGMG 538

Qy        540 GMGGMDGM 547

Db        539 GMGGMDGM 546


RESULT 5
US-10-369-493-10295
; Sequence 10295, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10295
;   LENGTH: 542
;   TYPE: PRT
;   ORGANISM: Cytophaga hutchinsonii
US-10-369-493-10295

  Query Match           67.8%;  Score 1835.5;  DB 12;  Length 542;
  Best Local Similarity  67.6%;  Pred. No. 1.7e-136;
  Matches  365;  Conservative   77;  Mismatches   97;  Indels    1;  Gaps    1;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62

Db          2 AKQILFDRDARERLKKGVDALADAVKVTLGPKGRNVILDKKFGSPAITKDGVSVAKDIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122

Db         62 KDAIENMGAQLVKEVASKTADQAGDGTTTATVLTQAIFNAGLKNVAAGANPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182

Db        122 AVSAIIADLKVQSKKISNSNEIAQVATISANNDHEIGKMIAHAMDKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242

Db        182 TETEVKTVEGMQFDRGYLSPYFVTNTDKMEVELDRPYILIYDKKVSSMKELLPILEQVVQ 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302

Db        242 SGRPLLIIAEDVDGEALATLVVNKIRGALKIAAVKAPGFGDRRKAMLEDIAILTGGTVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362

Db        302 EESGYKLENATLEYLGTAEKINIDKDNTTVVNGSGLKDNIVARVNQIKSQMENTTSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
```

```
Db        362 EKLQERLAKLSGGVAIMYIGAATEVEMKEKKDRVDDALHATRAAVEEGIVPGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482

Db        422 GAALDNVAFHNEDEKTGIQIIRTAIESPLRAIVYNAGLEGSVIVQKVKEGTGDYGYNARE 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDM-PMPGGGMGGM 541

Db        482 DRYEAMIAAGIIDPTKVTRLALENAASVASLLLTTECVVADEPEEKSAMPPMGGGGMGGM 541


RESULT 6
US-10-369-493-9287
; Sequence 9287, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9287
;   LENGTH: 541
;   TYPE: PRT
;   ORGANISM: Xylella fastidiosa
US-10-369-493-9287

  Query Match             67.7%;  Score 1834;  DB 12;  Length 541;
  Best Local Similarity   66.3%;  Pred. No. 2.2e-136;
  Matches  360;  Conservative   80;  Mismatches  101;  Indels    2;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKEIIFSEKARSRMVYGVNLLANAVKATLGPKGRHVVLDKSFGSPIITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELADKFENMGAQMLKEVASKTNDHAGDGTTTATVLAQALIREGCKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DKAVIAAVTELKKISKPTSDDKAIAQVATISANSDESIGNIIAEAMKKVGKEGVITIEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 TTLENELDVVEGMQFDRGYSSPYFINNQQSQIVELDNPYILLHDKKISSVRDLLTVLDAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 AKESKPLLIVAEEVEGEALATLVVNNIRGIIKVCAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEEVGLSLEKATTSHLGKAKKVRVSKENTTIIDGIGDNDAINGRVKQIKTQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVDDALLATRAAVEEGVIPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Db        421 RAITAISNLKGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGKDNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540

Db        481 ATGEFGDMVKLGILDPTKVARTALQNAASIAGLMITTEAMVAE--APKKDEPTPPAAGGG 538

Qy        541 MGG 543

Db        539 MGG 541


RESULT 7
US-10-369-493-15468
; Sequence 15468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15468
;   LENGTH: 544
;   TYPE: PRT
;   ORGANISM: Xanthomonas campestris
US-10-369-493-15468

  Query Match             67.7%;  Score 1833;  DB 12;  Length 544;
  Best Local Similarity   66.8%;  Pred. No. 2.7e-136;
  Matches  365;  Conservative   82;  Mismatches   95;  Indels    4;  Gaps    3;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKDIRFGEDARTRMVRGVNVLANAVKATLGPKGRNVVLEKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELADKFENMGAQMVKEVASRTNDNAGDGTTTATVLAQALIREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DQAVKAAVIELKNISKPTTDDKAIAQVGTISANSDESIGNIIAEAMQKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 SGLENELDVVEGMQFDRGYLSPYFINNQQSQSADLDDPFILLHDKKISNVRDLLPVLEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 AKAGKPLLIVAEEVEGEALATLVVNTIRGIVKVVAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEEVGLALEKATIKDLGRAKKVQVSKENTTIIDGAGDSATIEARVGQIKTQIEDTSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERVAKLAGGVAVIKVGASTEIEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Db        421 RALVAVGNLTGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGTGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MP-GGGM 538

Db        481 ANGEFGDMVEFGILDPTKVTRSALQNAASIAGLMITTEAMVAD--APKKDEPAMPAGGGM 538

Qy        539 GGMGGM 544

Db
```

```
Db        539 GGMGGM 544


RESULT 8
US-10-369-493-16217
; Sequence 16217, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16217
;   LENGTH: 544
;   TYPE: PRT
;   ORGANISM: Xanthomonas campestris
US-10-369-493-16217

  Query Match             67.7%;  Score 1833;  DB 12;  Length 544;
  Best Local Similarity   66.8%;  Pred. No. 2.7e-136;
  Matches  365;  Conservative   82;  Mismatches   95;  Indels    4;  Gaps    3;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:|  |    || :: |||: |||||| ||||||||||:|||||:| ||||||||||||
Db          1 MAAKDIRFGEDARTRMVRGVNVLANAVKATLGPKGRNVVLEKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||||||| :|:| ||||||||||||||: ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMVKEVASRTNDNAGDGTTTATVLAQALIREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |:|| |   || :|:||| | | |||||||||||| :||||||||| |||| |||||||
Db        121 DQAVKAAVIELKNISKPTTDDKAIAQVGTISANSDESIGNIIAEAMQKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  |||||||:|||||||||||: | :    :||:|:|| ::|||:::::|:|:|| |
Db        181 SGLENELDVVEGMQFDRGYLSPYFINNQQSQSADLDDPFILLHDKKISNVRDLLPVLEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :||||:||:|||||||||||| :|| ::|||||||||:|||||||:|:||||
Db        241 AKAGKPLLIVAEEVEGEALATLVVNTIRGIVKVVAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  ::  || ||:| : |||||:|||| |  |:||| ||: |||:|||||
Db        301 ISEEVGLALEKATIKDLGRAKKVQVSKENTTIIDGAGDSATIEARVGQIKTQIEDTSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||:|| |||||| ||||||:|||||||||:||||| | |
Db        361 DREKLQERVAKLAGGVAVIKVGASTEIEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |::  : ::  |::|:  |: |  |::| |||:| |||| | |::: ||:|    :|:||
Db        421 RALVAVGNLTGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGTGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MP-GGGM 538
              |:||: |::: |::|| |||| |||||||:| |::|||  :|:   |||| | || ||||
Db        481 ANGEFGDMVEFGILDPTKVTRSALQNAASIAGLMITTEAMVAD--APKKDEPAMPAGGGM 538

Qy        539 GGMGGM 544
              ||||||
Db        539 GGMGGM 544


RESULT 9
```

```
US-10-369-493-15265
; Sequence 15265, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15265
;   LENGTH: 543
;   TYPE: PRT
;   ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15265

  Query Match             67.5%;  Score 1828.5;  DB 12;  Length 543;
  Best Local Similarity   68.9%;  Pred. No. 6.1e-136;
  Matches  374;  Conservative   70;  Mismatches   96;  Indels    3;  Gaps    3;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||:   | ||||: :||| ||:|||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKPGASAREKMLKGVDILADAVKVTLGPKGRNVVIDKSFGAPPITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||:|:||| ||:|||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|     |      :|:|||||||| :  ||  |||||  :||  ||||||||
Db        121 DLAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||||||||| :|:: ||| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGTV 300
              :  :||:|||||||||  ||||||||||  |:: |||||||:|||||||||||||||||
Db        241 VQTGKPLVIIAEDVEGERLATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || :|:| | ||||||||||:  ||: || ||:||||||:|||
Db        301 ISEDLGIKLESVTLDMLGKSKKVSISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:||||||::||||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479
              ||   :  :|  :||: ||:||:|::|:  :|||| ||| | |||| |: :  :| :|:|
Db        421 RSSTKI-TVKGVNDDQEAGINIVRKALQSLVRQIAENAGDEASIVVGKILDKNEDNYGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538
              | :||| |||  |::|| || | ||||||||||:|:||| ||| |: :  || ||||||
Db        480 AQTGEYGDLIALGIVDPVKVVRTALQNAASVASVLITTEAMIAELPKKESAMPQMPGGGM 539

Qy        539 GGM 541
              |||
Db        540 GGM 542


RESULT 10
US-10-369-493-11922
; Sequence 11922, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
```

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11922
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-11922

Query Match 67.4%; Score 1826; DB 12; Length 542;
Best Local Similarity 67.8%; Pred. No. 9.5e-136;
Matches 366; Conservative 80; Mismatches 92; Indels 2; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||:  | : ||||: |||: ||:|||||||||||||||:||||:| ||||||:|||||
Db          1 MAAKEVKFHSDAREKMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||||:||| ||||:|||||||||||||||| :|| | ||:| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTSDLAGDGTTTATVLAQAIVKEGAKAVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |: :||    :      |||||||||| |  ||  :|||| |||  ||||||||
Db        121 DKAVEAIVQELKTNARKVTRNDEIAQVGTISANGDAEIGRFLAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||:|| :||  ||: ||:| :|||:::::: :||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFITNQDKMRVELEEPYVLIHEKKLSNLQALLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               : ::|||||||||||||||||||||||| :| ||||||||:|||||||||||||||| |
Db        241 VQSSKPLLIIAEDVEGEALATLVVNKLRGGLKVAAVKAPGFGDRRKAMLEDIAILTGGTA 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVEQIRAQIEETSSDY 360
              | || |||||||:|  || ||:|||:||||||||||  ::|: || ||::|||||:|||
Db        301 ISEDLGIKLENVTLEMLGRAKKVVIEKENTTIVDGAGSKDEIQGRVSQIKSQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:|||||||||| |||||| || :|| |:||:||||:||::|||||||||::|||| | :
Db        361 DKEKLQERLAKLAGGVAVIRVGGSTEVEVKERKDRVDDAMHATRAAVEEGVLPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVRE-PKDGFGFN 479
              |: | || ::  ::|:  |: |:|||:| |:|||| ||| ||||:| |:|| |:  |:|:|
Db        421 RAAKALDSVQAENEDQKHGIEIVRRAIEAPVRQIAENAGAEGSIIVGKLREKPEFGWGWN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538
              | :  : ||   ||||| || | |||:||||| ||:||| :||||: :  :| || |||
Db        481 AQTNAFGDLYNEGVIDPVKVVRTALQDAASVAGLLVTTEAMVAEKPKKEPAVPAMPAGGM 540
```

RESULT 11
US-10-369-493-12330
; Sequence 12330, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12330
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12330

Query Match 66.9%; Score 1812.5; DB 12; Length 551;
Best Local Similarity 67.8%; Pred. No. 1.1e-134;
Matches 375; Conservative 71; Mismatches 96; Indels 11; Gaps 5;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   |||:: |||: ||:|||||||||||||||:||||:| ||||||:|||||
Db          1 MAAKDVKFSRDARERMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQSIVQEGHKAVAAGMNPMDLKRGI 120
              |||||||||||||:||| ||:|||||||||||||||:| :|| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQSIVQEGHKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  |   |    | :  :|:|||||||| | ::| :|||| ||| ||||||||
Db        121 DLAVSEVVAALGKAAKKIKTSEEVAQVGTISANGDESVGKMIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||| :||| :|:: ||| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNADKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               : ::|||||:|||||||||||||||| |:: ||||||||:|||||||||||||||:
Db        241 VQTSKPLLIISEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || ||||||| |: || ||:| | ||||||||||||  :|: || ||: |||||:|||
Db        301 ISEDLGIKLENVGLNMLGRAKKVSISKENTTIVDGAGKKAEIQGRVAQIKQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVAGGGVALL 420
              |:|||||||||| |||||| || ||| |:||||||:|||||||||||||||| ||| | :
Db        361 DKEKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGVALL 420

Qy        421 RSIKVLDDIKP--ADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDG-FG 477
              |:     :||   |: |: ||:||:||:|: | ||||:||| | |||  |: | |    ||
Db        421 RASA---NIKAAGANADQAAGINIVRRALQAPARQIASNAGAEASIVAGKILENKGATFG 477

Qy        478 FNAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKK---DMP-- 532
              :|| :||| |:|   |::|| || | |||:||||| ||:||| |||  |: :     ||
Db        478 YNAQTGEYGDMIAMGIVDPVKVVRTALQDAASVAGLLVTTEAMIAEAPKKESAGGGMPGG 537

Qy        533 MPGGGMGGMGGMD 545
              | ||||||||||
Db        538 MGGGGMGGMGGMD 550
```

RESULT 12
US-10-369-493-15838
; Sequence 15838, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15838
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15838

Query Match 66.9%; Score 1811; DB 12; Length 540;
Best Local Similarity 66.6%; Pred. No. 1.5e-134;
Matches 361; Conservative 82; Mismatches 95; Indels 4; Gaps 3;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:| |   || :: |||: |||||| ||||||||||:|||||:| ||||||||||||
Db          1 MAAKDIRFGEDARTRMVRGVNVLANAVKATLGPKGRNVVLEKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||||||| :|:| |||||||||||||: ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMVKEVASRTNDNAGDGTTTATVLAQALIREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |:|| |   || :|:||| | | |||||||||||| :|||||||| |||| |||||||
Db        121 DQAVKAAVIELKNISKPTTDDKAIAQVGTISANSDESIGNIIAEAMQKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  |||||||:|||||||||: | :    :||:|:|| ::|||:::::|:||:|| |
Db        181 SGLENELDVVEGMQFDRGYLSPYFINNQQSQSADLDDPFILLHDKKISNVRDLLPVLEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :||||:||:|||||||||||| :|| ::|||||||||:|||||||:|:|||
Db        241 AKAGKPLLIVAEEVEGEALATLVVNTIRGIVKVVAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  ::  || ||:| : |||||:|||| |  |:||| ||: |||:|||||
Db        301 ISEEVGLALEKATIKDLGRAKKVQVSKENTTIIDGAGDSATIEARVGQIKTQIEDTSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||:|| |||||| ||||||:|||||||||:|||||  | |
Db        361 DREKLQERVAKLAGGVAVIKVGASTEIEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |::  : ::  |::|:  |: |  |::| |||:| |||| | |::: ||:|    :|:||
Db        421 RALVAVGNLTGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGTGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MP-GGGM 538
              |:||: |::: |::|| |||| |||||||:| |::|||  :|:   |||| | || ||||
Db        481 ANGEFGDMVEFGILDPTKVTRSALQNAASIAGLMITTEAMVAD--APKKDEPAMPAGGGM 538

Qy        539 GG 540
              ||
Db        539 GG 540
```

RESULT 13
US-10-369-493-12170
; Sequence 12170, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12170
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12170

Query Match 66.8%; Score 1810; DB 12; Length 552;
Best Local Similarity 67.6%; Pred. No. 1.8e-134;
Matches 373; Conservative 72; Mismatches 99; Indels 8; Gaps 4;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   |||:: ||||: ||:|||||||||||||:|||||:| ||||||:|||||
Db          1 MAAKDVKFSRDARERMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQSIVQEGHKAVAAGMNPMDLKRGI 120
              |||||||||||||:||| ||:|||||||||||||:| :|| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQSIVQEGHKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | || |   |        :|:|||||: | | ::| :||||| ||| ||||||||
Db        121 DLAVSDVVWTLIKNATKIKTSEEVAQVGTIAGNGDESVGKMIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||| :||| :|:: ||| :|||:::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNADKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              : ::|||||:|||||||||||||||| |:: ||||||||:|||||||||||||:
Db        241 VQTSKPLLIISEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || ||||||| |  || ||:| | |||||||||||| |:|: || ||: |||||:|||
Db        301 ISEDLGIKLENVGLDMLGRAKKVSISKENTTIVDGAGKKEEIQGRVAQIKQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:|||||||||| |||||| || ||| |:|||||||:|||||||||||||||||||| | :
Db        361 DKEKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDG-FGFN 479
              |:   :: :  |: |: ||::|:||:|: | |||||||| | |||  |: | |   ||||
Db        421 RASLSINAV-GANSDQTAGISIVRRALQAPARQIAANAGAEASIVAGKILENKGATFGFN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKK----DMP--M 533
              | :||| |:|  |::|| || | |||:||||| ||:||| |||  |: :       ||  |
Db        480 AQTGEYGDMIAMGIVDPVKVVRTALQDAASVAGLLVTTEAMIAEAPKKESAGGGGMPGGM 539

Qy        534 PGGGMGGMGGMD 545
              ||||||||||||
Db        540 GGGGMGGMGGMD 551
```

RESULT 14
US-10-369-493-167
; Sequence 167, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 167
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Aquifex aeolicus
US-10-369-493-167

Query Match 66.0%; Score 1786; DB 12; Length 545;

```
  Best Local Similarity   64.9%;  Pred. No. 1.4e-132;
  Matches  352;  Conservative   94;  Mismatches   86;  Indels   10;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:| |::: :|| ||  |||||||||||||||||| |:: |::|:||:|||||:|||||
Db          1 MAAKAIIYNEEARAKLKAGVDKLANAVKVTLGPKGREVILGKNWGTPVVTKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||:|||:|||||  ||:|:|||||||||||||||: ||::: |:| | | :||||
Db         61 ELKDKFENIGAQLVKEVASKTADVAGDGTTTATVLAQAIFHEGLRVAASGANVMEVKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||  : :||  ::|  :::||| || |||||:|  || |||:|| :||| ||||||||:
Db        121 DKAVKKIVEELKKLSKDVKERKEIEQVATISANNDPEIGKIIADAMEEVGKDGVITVEES 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||||:||:||:||||||||||||:|||| | |:|||||  |||||::|::|||||||
Db        181 KSAETTLEVVKGMQFDRGYLSPYFVTDPEKMECVLENPYILIYEKKITNVKELLPILEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  ||||:|||||||||||||||| ::| |:  |||||||:|||  | |||:||||:|
Db        241 VRSGRPLLVIAEDVEGEALATLVVNHIKGVLKACAVKAPGFGQRRKDYLGDIAVLTGGQA 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || |::||:|||:|||: | |  | ||||::||: ||:||:|||
Db        301 ITEDLGIKLESVTLDMLGQAEKVVVDKEHTTIIGGKGDPEQIKARIEQIKRQIQETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| ||||:| |||||| |:|||| |||||::||:||||||||||| | |
Db        361 DREKLQERLAKLSGGVAIIRVGAATEAELKEKKYRVEDAVHATKAAVEEGIVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKV----REPKDGF 476
              |: : |:|:|  : |:  |::||:::: ||:||| ||||:||:|:|||    :|   :
Db        421 RASEALEDLKGDNHDQQLGIDIIKKAVRTPLKQIAYNAGYDGSVVLEKVIELGKEKGVSW 480

Qy        477 GFNAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKK------D 530
              |||||:||| |: :||:||| || | |::|||||  :|| |  ||: || ||      |
Db        481 GFNAATGEYVDMYEAGIIDPTKVVRTAIENAASVAGTMLTAEALIADLPEEKKKDITPTD 540

Qy        531 MP 532
              ||
Db        541 MP 542


RESULT 15
US-10-369-493-17132
; Sequence 17132, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17132
;   LENGTH: 544
;   TYPE: PRT
;   ORGANISM: Bacillus halodurans
US-10-369-493-17132

  Query Match             65.7%;  Score 1780.5;  DB 12;  Length 544;
  Best Local Similarity   64.8%;  Pred. No. 3.8e-132;
  Matches  355;  Conservative   89;  Mismatches   95;  Indels    9;  Gaps    4;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|:| |   ||  : |||||||:||||||||||||:|| ||||:|| |||::||||||
Db          2 AKDIKFSEDARRSMLRGVDKLADAVKVTLGPKGRNVVLEKKFGSPLITNDGVTIAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              || ||||||::| ||| ||:|||||||||||||||: |||:| | :| ||| |::||:|
Db         62 EDAFENMGAKLVAEVASKTNDIAGDGTTTATVLAQAMIREGLKNVTSGANPMVIRKGIEK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              |     :||| |:||   :  |||| || ::|  :| |||||:|  :|||::::::||:|||| :
Db        122 ATQVAVEELSKISKPIEGKDSIAQVAAIS-SADDEVGKIIAEAMERVGNDGVITIEESKG 180

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              | |:|||||:||||| ||| ||: :||   |||||:|  :|||::::::||:|||| :
Db        181 FSTELEVVEGMQFDRGYASPYMVTDSDKMEAVLDNPYVLITDKKISNIQEVLPVLEQVVQ 240

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :|:|||||||||||||||||||    |||||||||:|||||||||||||||| |
Db        241 QGKPILIIAEDVEGEALATLVVNKLRGTFNAVAVKAPGFGDRRKAMLEDIAILTGGEVIT 300

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              || |: |:: ::: || | :||: |||||||:|||:|: | ||| ||:||||||:||:|:
Db        301 EDLGLDLKSANITQLGRASKVVVTKENTTIVEGAGESDKIAARVNQIKAQIEETTSDFDK 360

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| ||||||: ||||||||||:| |:|||||:|||||||||| ||||| |
Db        361 EKLQERLAKLAGGVAVLKVGAATETEMKERKLRIEDALNSTRAAVEEGIVAGGGTALVNV 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              || :  |  |: ||  |:||: |:|||:|||| ||| |||::||:::: : |||||||:
Db        421 IKAVSSI-GAEGDEATGVNIVLRALEEPVRQIAHNAGLEGSVIVERLKKEEAGFGFNAAT 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGG---MG 539
              ||: ::::||::|| |||| |||:||||::: |||| ||:||| :     ||| ||
Db        480 GEWVNMVEAGIVDPTKVTRSALQHAASVSAMFLTTEAVIADKPEENEG----GGGMPDMG 535

Qy        540 GMGGMDGM 547
              ||||| ||
Db        536 GMGGMGGM 543


Search completed: January 28, 2004, 13:22:58
Job time : 39 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:11:55 ; Search time 21 Seconds
(without alignments)
2509.545 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	68.9	551	2	JC2562	chaperonin groELx
2	1864	68.8	545	2	C95311	groEL2 chaperonin
3	1863	68.8	545	2	JN0509	heat shock protein
4	1862.5	68.8	547	2	F82783	60kDa chaperonin X
5	1852.5	68.4	546	2	B47073	chaperonin GroEL -
6	1845	68.1	550	2	A41468	60K heat shock pro
7	1844.5	68.1	546	2	S65596	heat shock protein
8	1843.5	68.1	547	2	I40331	Cpn60 protein (Gro
9	1843.5	68.1	552	2	S39765	chaperonin 60 - Co
10	1841	68.0	542	2	F95967	probable heat shoc
11	1840.5	68.0	544	2	AD2660	60 KDA chaperonin
12	1840.5	68.0	544	2	B97442	60K chaperonin (pr
13	1828.5	67.5	544	2	S23918	groEL protein - Ag
14	1825.5	67.4	547	2	B83098	GroEL protein PA43
15	1823	67.3	546	2	S22347	groEL - Brucella a
16	1823	67.3	550	2	S35309	heat shock protein
17	1822	67.3	546	2	I40342	heat shock protein
18	1817.5	67.1	544	2	B36917	heat shock protein
19	1816.5	67.1	541	2	JN0512	heat shock protein
20	1814.5	67.0	544	2	H81964	chaperonin 60kD su
21	1814.5	67.0	549	2	JC2564	heat shock protein
22	1814	67.0	546	2	AG3640	60K chaperonin gro
23	1812.5	66.9	545	2	S47530	chaperonin groEL -
24	1810.5	66.9	544	2	S37039	groEL protein - Ba
25	1808.5	66.8	544	2	C81021	chaperonin, 60 kDa
26	1805.5	66.7	547	2	B43606	heat shock protein
27	1802.5	66.6	544	2	S61300	heat shock protein
28	1799	66.4	546	2	B54539	heat shock protein
29	1794.5	66.3	544	2	S61302	heat shock protein
30	1794	66.2	546	2	S35311	heat shock cognate
31	1791.5	66.2	547	2	B87334	chaperonin, 60 kDa
32	1791	66.1	541	2	S61301	heat shock protein
33	1789.5	66.1	541	2	JN0511	heat shock protein
34	1788.5	66.0	544	2	S61303	heat shock protein
35	1786	66.0	545	2	C70489	GroEL - Aquifex ae
36	1780.5	65.7	544	2	JC6063	chaperonin groEL -
37	1780.5	65.7	544	2	B83720	class I heat-shock
38	1774.5	65.5	545	2	S51563	heat shock protein
39	1773	65.5	544	1	B43827	chaperonin groEL -
40	1765	65.2	546	2	S34938	heat shock protein
41	1761.5	65.0	544	2	B82048	chaperonin, 60 Kd
42	1761	65.0	548	2	AG0043	60 kDa chaperonin
43	1759	65.0	548	2	C64076	chaperonin groEL -
44	1755.5	64.8	539	2	B49855	heat shock protein
45	1755	64.8	545	2	G81328	60 kD chaperonin (

ALIGNMENTS

RESULT 1
JC2562
chaperonin groELx protein - Amoeba proteus
C;Species: Amoeba proteus
C;Date: 17-May-1995 #sequence_revision 14-Jul-1995 #text_change 26-Aug-1999
C;Accession: JC2562
R;Ahn, T.I.; Lim, S.T.; Leeu, H.K.; Lee, J.E.; Jeon, K.W.
Gene 148[128], 43-49, 1994
A;Title: A novel strong promoter of the groEx operon of symbiotic bacteria in Amoeba prot
A;Reference number: JC2561
A;Note: due to a typographical error the volume number 148 appears as 128
A;Accession: JC2562
A;Molecule type: DNA
A;Residues: 1-551 <AHN>
A;Cross-references: GB:M86549; NID:g155400; PIDN:AAC09381.1; PID:g155402
C;Comment: This protein is involved in the assembly of oligomeric protein complexes, and
C;Genetics:
A;Gene: groELx
C;Superfamily: chaperonin groEL
C;Keywords: molecular chaperone

Query Match 68.9%; Score 1866; DB 2; Length 551;
Best Local Similarity 67.9%; Pred. No. 2.2e-83;
Matches 373; Conservative 74; Mismatches 96; Indels 6; Gaps 2;

```
Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   ||:::  ||: ||: || |:|| ||||||:|:|||:| :||||||||||||
Db          2 AKELRFGDDARQQMLAGVNALADRVKATMGPSGRNVVLERSFGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              |::|:|||||||||||| |||| ||||||||||||::|  || | |||| ||| :||||||
Db         62 ENRFKNMGAQMVKEVAAKTSDTAGDGTTTATVLARSIVVEGHKAVAAGMNPMDLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              || |:|||  ::|| :| | ||||||||||||  ||:|||||| |||| ||||||:  |
Db        122 AVTAITKELQKMSKPCKDGKAIAQVGTISANSDQAIGSIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:|||| | : |  ||:|:|:||  :||| :::||| :|| |||
Db        182 LENELSVVEGMQFDRGYISPYFINNQQNMSAELEHPFILLVDKKIATIRDMLSVLEAVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              |||||:|||||||||||||| :|| ::| ||||||||:||||||:|||||| |: |
Db        242 SGRPLLIVAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTNGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  || |||||||:|: ||||||:|| ||: :| ||: |||||:|||||||||
Db        302 EEIGTSLETASLESLGTAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETSSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              ||||||:||| |||||| |||||| |||||| ||||||:|||||||||| ||| | :|:
Db        362 EKLQERVAKLAGGVAVIKVGAATEIEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421
```

```
Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              |||| :|  : |:  |:||:||::| ||||| |||||| |::| || | || ||||||:
Db        422 QKVLDGLKGDNADQDMGINILRRAIESPLRQIVANAGYESSVIVNKVAEHKDNFGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG----GGM 538
              |:| |::: |::|| |||| |||||||| ||:||||| :|:   |||  | |    |||
Db        482 GQYGDMVEMGILDPTKVTRTALQNAASVRSLMLTTECMVADL--PKKDEGMAGAGDMGGM 539

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        540 GGMGGMGGM 548
```

RESULT 2
C95311
groEL2 chaperonin [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: C95311
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: C95311
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-545 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK65053.1; PID:g14523485; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: groEL2
A;Genome: plasmid
C;Superfamily: chaperonin groEL

  Query Match 68.8%; Score 1864; DB 2; Length 545;
  Best Local Similarity 70.5%; Pred. No. 2.7e-83;
  Matches 387; Conservative 59; Mismatches 97; Indels 6; Gaps 4;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||||: |||| ||:|||||||||||||||:||||:| ||||||:|||||
Db          1 MAAKEVKFGRSAREKMLRGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||:||| ||:|||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|    |       |:|||||||| :  ||  ||||| ||| ||||||||
Db        121 DLAVAEVVKDLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:||||||||||||||||| :|:: :|| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAFILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  :|||||||||||||||||||||||| |:: ||||||||:||||||||||||||||
Db        241 VQTGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || ||:| | ||||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSITKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:||||||::|||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGF 478
              | |:|:   :|  :||: ||:||:||:|: | |||  ||| | |||| |: |   | ||:
Db        421 RSSVKI--TVKGENDDQDAGVNIVRRALQSPARQIVENAGDEASIVVGKILEKNTDDFGY 478

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGM 538
              || :||| |:|  |:||| || | |||:||||||||:|||  |||   |||| |   |||
Db        479 NAQTGEYGDMIAMGIIDPVKVVRTALQDAASVASLLITTEAMIAEL--PKKDAPAMPGGM 536

Qy        539 GGMGGMDGM 547
              ||||||| |
Db        537 GGMGGMDMM 545
```

RESULT 3
JN0509
heat shock protein groEL (clone Rhz A) - Rhizobium meliloti
N;Alternate names: chaperonin groEL protein
C;Species: Rhizobium meliloti
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C;Accession: JN0509
R;Rusanganwa, E.; Gupta, R.S.
Gene 126, 67-75, 1993
A;Title: Cloning and characterization of multiple groEL chaperonin-encoding genes in Rhi
A;Reference number: JN0509; MUID:93231539; PMID:8097179
A;Accession: JN0509
A;Molecule type: DNA
A;Residues: 1-545 <RUS>
A;Cross-references: GB:M94192; NID:g152233; PIDN:AAA26285.1; PID:g152235
C;Comment: This protein plays a role in protein folding and in the extracellular transpo
C;Genetics:
A;Gene: groEL
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone; stress-induced protein

  Query Match 68.8%; Score 1863; DB 2; Length 545;
  Best Local Similarity 70.5%; Pred. No. 3e-83;
  Matches 387; Conservative 58; Mismatches 98; Indels 6; Gaps 4;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   |||: |||| ||:|||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFGRSGREKMLRGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVREVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||:||| ||:|||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|    |       |:|||||||| :  ||  ||||| ||| ||||||||
Db        121 DLAVAEVVKDLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:||||||||||||||||| :|:: :|| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAFILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  :|||||||||||||||||||||||| |:: ||||||||:||||||||||||||||
Db        241 VQTGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || ||:| | ||||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSITKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:||||||::|||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGF 478
              | |:|:   :|  :||: ||:||:||:|: | |||  ||| | |||| |: |   | ||:
Db        421 RSSVKI--TVKGENDDQDAGVNIVRRALQSPARQIVENAGDEASIVVGKILEKNTDDFGY 478
```

```
Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGM 538
              || :||| |:|  |:||| || | |||:||||||||:|||  |||   |||| |   |||
Db        479 NAQTGEYGDMIAMGIIDPVKVVRTALQDAASVASLLITTEAMIAEL--PKKDAPAMPGGM 536

Qy        539 GGMGGMDGM 547
              |||||||
Db        537 GGMGGMDMM 545


RESULT 4
F82783
60kDa chaperonin XF0615 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: F82783
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <SIM>
A;Cross-references: GB:AE003907; GB:AE003849; NID:g9105484; PIDN:AAF83425.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0615
C;Superfamily: chaperonin groEL

  Query Match             68.8%;  Score 1862.5;  DB 2;  Length 547;
  Best Local Similarity   66.8%;  Pred. No. 3.2e-83;
  Matches  366;  Conservative   80;  Mismatches   97;  Indels    5;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|||:|  |||  ::  ||: |||||| |||||||:||::||||||:|||||||||||
Db          1 MAAKEIIFSEKARSRMVHGVNLLANAVKATLGPKGRHVVLDKSFGSPIITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||:|||| ||:| |||||||||||||: ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMLKEVASKTNDHAGDGTTTATVLAQALIREGCKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||:|   ||  |:||| | | |||| |||||| :||||||| |||| ||||:||
Db        121 DKAVIAAVTELKKISKPTSDDKAIAQVATISANSDESIGNIIAEAMKKVGKEGVITIEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||  |||||||:||||| |||| | :  : |||||||| ::|||:|::|:| :|: |
Db        181 TTLENELDVVEGMQFDRGYSSPYFINNQQSQIVELDNPYILLHDKKISSVRDLLTVLDAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || ::|||:||:|||||||||||| :|| ::| ||||||||:||||||||:|:||||
Db        241 AKESKPLLIVAEEVEGEALATLVVNNIRGIIKVCAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  : | || ||:| : ||||||:|| | :: |  |||||: |||||:|||
Db        301 ISEEVGLSLEKATTSHLGKAKKVRVSKENTTIIDGIGDNDAINGRVKQIKTQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||:||| |||||||||::|||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVDDALLATRAAVEEGVIPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:|  : ::| |::|:  |: |  |::| |||:| |||| | |::: ||:| || ||:||
Db        421 RAITAISNLKGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGKDNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMP---GGG 537
              |:||: |::  |::|| |||| |||||||:| |::|||  :||   |||| | |   |||
Db        481 ATGEFGDMVNLGILDPTKVTRSALQNAASIAGLMITTEAMVAE--APKKDEPTPPAAGGG 538

Qy        538 MGGMGGMD 545
              ||||||||
Db        539 MGGMGGMD 546


RESULT 5
B47073
chaperonin GroEL - Chromatium vinosum
C;Species: Chromatium vinosum
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C;Accession: B47073
R;Ferreyra, R.G.; Soncini, F.C.; Viale, A.M.
J. Bacteriol. 175, 1514-1523, 1993
A;Title: Cloning, characterization, and functional expression in Escherichia coli of chap
A;Reference number: A47073; MUID:93186721; PMID:8444812
A;Accession: B47073
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-546 <FER>
A;Cross-references: GB:M99443; NID:g145007; PIDN:AAA23319.1; PID:g145009
A;Note: sequence extracted from NCBI backbone (NCBIN:126968, NCBIP:126970)
C;Superfamily: chaperonin groEL
C;Keywords: molecular chaperone

  Query Match             68.4%;  Score 1852.5;  DB 2;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 9.7e-83;
  Matches  374;  Conservative   70;  Mismatches  101;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              |::|:: |   || ::  ||: ||||||||||||||||:|||||:| :||||||||||
Db          1 MSAKDVKFGGDARVRMMEGVNILANAVKVTLGPKGRNVVLEKSFGAPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||||||| ||||||||||||||||||: |||:| |||| ||| :|||:
Db         61 ELKDKFENMGAQMVKEVASKTSDIAGDGTTTATVLAQAMVREGLKAVAAGMNPMDLKRGM 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDDSIGTIIAEAMEKVGKEGVITVEDG 180
              |||| | |:||  ::||       |||||||||||| :|| |||||| |||| ||||||:
Db        121 DKAVEAATEELKKLSKPCPRPMAIAQVGTISANSDDSIGTIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |:  |||||||:||||||||||: | :     ||| |||| :|||:::::|:||:|| |
Db        181 TSLQNELDVVEGMQFDRGYLSPYFINNQQSQSAELDAPYILLYDKKISNIRDLLPVLEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||||||| :|| ::| ||||||||:||||||:|||||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTIRGIVKVCAVKAPGFGDRRKAMLQDIAILTGATV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: ||||||| : |: |||:||:|   ||||| :|||||:|||||||
Db        301 ISEEVGLSLEKATLTDLGTAKRVQVGKDETTIIDGSGSEIDIKARCEQIRAQVEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||| |||||| |||||| |||||| |||||:|||||||||||||| | |
Db        361 DREKLQERLAKLAGGVAVIKVGAATEIEMKEKKARVEDALHATRAAVEEGIVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:|  : |:| |: |:  |: | ||::||||||| |||| | |::: || |    ||:||
Db        421 RAIAAVKDLKGANHDQDVGIAIARRAMEEPLRQIVANAGEEPSVILHKVAEGTGNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGMG 539
```

```
Db        481 ANGEYGDMVEMGILDPTKVTRSALQNSCSVAGLMITTEAMIAD--EPKDDAPAMPGGGMG 538

Qy        540 GMGGMDGM 547
              |||| |
Db        539 DMGGMGMM 546


RESULT 6
A41468
60K heat shock protein htpB - Legionella pneumophila
C;Species: Legionella pneumophila
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C;Accession: A41468; A41472
R;Hoffman, P.S.; Houston, L.; Butler, C.A.
Infect. Immun. 58, 3380-3387, 1990
A;Title: Legionella pneumophila htpAB heat shock operon: nucleotide sequence and express
A;Reference number: A41468; MUID:90382960; PMID:2205580
A;Accession: A41468
A;Molecule type: DNA
A;Residues: 1-550 <HOF>
A;Cross-references: GB:M31918
R;Sampson, J.S.; O'Connor, S.P.; Holloway, B.P.; Plikaytis, B.B.; Carlone, G.M.; Mayer,
Infect. Immun. 58, 3154-3157, 1990
A;Title: Nucleotide sequence of htpB, the Legionella pneumophila gene encoding the 58-ki
A;Reference number: A41472; MUID:90354095; PMID:2117582
A;Accession: A41472
A;Molecule type: DNA
A;Residues: 3-188,'S',190-195,'LIAVH',202-347,'T',349-550 <SAM>
A;Cross-references: GB:M91673; GB:M35149; NID:g149689; PIDN:AAA25298.1; PID:g149690
C;Genetics:
A;Gene: htpB
C;Superfamily: chaperonin groEL

  Query Match             68.1%;  Score 1845;  DB 2;  Length 550;
  Best Local Similarity   67.3%;  Pred. No. 2.3e-82;
  Matches  367;  Conservative   80;  Mismatches   96;  Indels    2;  Gaps    1;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|: |   || ::  ||: ||:||:||:||:|||||:|||:|:| :|||||||||||||
Db          4 AKELRFGDDARLQMLAGVNALADAVQVTMGPRGRNVVLEKSYGAPTVTKDGVSVAKEIEF 63

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
             | :| |||||||||||| |||| |||||||||||||::|  || | |||| ||| :||||||
Db         64 EHRFMNMGAQMVKEVASKTSDTAGDGTTTATVLARSILVEGHKAVAAGMNPMDLKRGIDK 123

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||:||||:|  ::|| :| | |||||||||||||  || |||||| |||| ||||||:  |
Db        124 AVLAVTKKLQAMSKPCKDSKAIAQVGTISANSDEAIGAIIAEAMEKVGKEGVITVEDGNG 183

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              ||  | |||||:|||||:||||: | : | |||::|:||  :||::|::||  :|| |||
Db        184 LENELYVVEGMQFDRGYISPYFINNQQNMSCELEHPFILLVDKKVSSIREMLSVLEGVAK 243

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
                |||||||||||||||||||| :|| ::| |||||||:||||||:|||||| |: |
Db        244 SGRPLLIIAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTKGQVIS 303

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |  ||  :|  ||:|||:|: ||||||:|| ||: :| ||: |||||:||||:|||||
Db        304 EEIGKSLEGATLEDLGSAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETTSDYDR 363

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||:||| |||||| |||||| |||||| ||||||:|||||||||||| ||| | :|:
Db        364 EKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 423

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | || :|  :||:  |:|||:||::| |:|||  ||||| |:|| || | || :||||||:
Db        424 QKALDSLKGDNDDQNMGINILRRAIESPMRQIVTNAGYEASVVVNKVAEHKDNYGFNAAT 483

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||| |::: |::|| |||||||||||||:||||| :|:   |||:  :  | |||||
Db        484 GEYGDMVEMGILDPTKVTRMALQNAASVASLMLTTECMVADL--PKKEEGVGAGDMGGMG 541

Qy        543 GMDGM 547
              || ||
Db        542 GMGGM 546


RESULT 7
S65596
heat shock protein 60 - Rhizobium leguminosarum
N;Alternate names: chaperonin 60
C;Species: Rhizobium leguminosarum
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C;Accession: S65596
R;Wallington, E.J.; Lund, P.A.
Microbiology 140, 113-122, 1994
A;Title: Rhizobium leguminosarum contains multiple chaperonin (cpn60) genes.
A;Reference number: S65596; MUID:94214663; PMID:7909257
A;Accession: S65596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <WAL>
A;Cross-references: EMBL:L20775; NID:g387873; PIDN:AAA26246.1; PID:g387874
C;Genetics:
A;Gene: cpn60
C;Superfamily: chaperonin groEL
C;Keywords: heat shock; molecular chaperone

  Query Match             68.1%;  Score 1844.5;  DB 2;  Length 546;
  Best Local Similarity   69.8%;  Pred. No. 2.4e-82;
  Matches  384;  Conservative   61;  Mismatches   98;  Indels    7;  Gaps    5;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              |||||| |    ||: |||| ||:|||||||||||||:|:||||:| ||||||||||||
Db          1 MASKEIKFGRTGREKMLRGVDILADAVKVTLGPKGRNVIIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||:||| ||:|||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGNKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|    |     :|:|||||||| |  :|  |||| ||| ||||||||
Db        121 DLAVADVVKDLQAKAKKISTSEEVAQVGTISANGDKQVGLDIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||||||||: :|:: :|| :|||:::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMIADLEDVFILLHEKKLSNLQSMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  :||||:|||||||||||||||||| |:: |||||||:||| ||||||||||
Db        241 VQTGKPLLIVAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKRMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || ||:| | |||||||||:|   ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSISKENTTIVDGSGAKTDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:|||||::|||||||||:|||||||| |
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGIALA 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479
              ||   :  :| |:||: ||:||:||:|:  :|||| ||| | |||| || :  :| ||:|
Db        421 RSSTKI-TVKGANDDQEAGINIVRRALQSLVRQIAENAGDEASIVVGKVLDKNEDNFGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP--MPGGG 537
              | : || |:|  |::|| || | ||||||||||||:|||  |||   |||| |  || ||
Db        480 AQTSEYGDMIAMGIVDPLKVVRTALQNAASVASLLITTEAMIAEL--PKKDAPAGMP-GG 536

Qy        538 MGGMGGMDGM 547
              |||||||| |
Db        537 MGGMGGMDMM 546


RESULT 8
```

```
I40331
Cpn60 protein (GroEL) - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 26-Aug-1999
C;Accession: I40331
R;Fernandez, R.C.; Weiss, A.A.
Gene 158, 151-152, 1995
A;Title: Cloning and sequencing of the Bordetella pertussis cpn10/cpn60 (groESL) homolog
A;Reference number: I40330; MUID:95309719; PMID:7789805
A;Accession: I40331
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-547 <RES>
A;Cross-references: EMBL:U12277; NID:g968918; PIDN:AAA74967.1; PID:g968920
C;Genetics:
A;Gene: cpn60
C;Superfamily: chaperonin groEL

  Query Match             68.1%;  Score 1843.5;  DB 2;  Length 547;
  Best Local Similarity   67.4%;  Pred. No. 2.7e-82;
  Matches  370;  Conservative   74;  Mismatches   98;  Indels    7;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|::||  :|| :: |||: |||||| ||||||||||:|:|||:| :||||||||||
Db          1 MAAKQVLFADEARVRIVRGVNVLANAVKTTLGPKGRNVVLERSFGAPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||:|||:||:|| |||| |||||||||||||: :||:| |||| ||: :||||
Db         61 ELKDKFENIGAQLVKDVASKTSDNAGDGTTTATVLAQAVVQEGLKYVAAGFNPIDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  ::||    |||||:|||||| :|| |||:|| |||| ||||||:
Db        121 DKAVAAAVEELKKLSKPVTTSKEIAQVGSISANSDASIGQIIADAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | ||  |||||||:||||||||||: :||| |  ||:||:|  :||:::::|:||:||||
Db        181 KSLENELDVVEGMQFDRGYLSPYFINSPEKQVAALDDPYVLIYDKKVSNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || :|||||||||||||||||||| :|| |:  |||||||:|||||||||||||||
Db        241 AKSSRPLLIIAEDVEGEALATLVVNNIRGILKTTAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|  || |||: : |||||:||||  : |:||||||||||| :|||
Db        301 ISEETGMSLEKATLQDLGQAKRIEVAKENTTIIDGAGDGKSIEARVKQIRAQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||||||:|||||||||:||||| | :
Db        361 DREKLQERVAKLAGGVAVIRVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |: : :  :|    |: ||: :| |::||||| |  ||| | |:||  |   | :|:||
Db        421 RAKQAITGLKGDTADQNAGIKLILRAVEEPLRTIVTNAGDEASVVVNTVLNGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAE----KPEPKKDMPMPGG 536
              |:||| ||:: ||:|| |||| |||||||||||||| | |: |    ||     ||   |
Db        481 ATGEYGDLVEQGVLDPTKVTRTALQNAASVASLLLTAEAAVVELMENKPAAAPAMP---G 537

Qy        537 GMGGMGGMD 545
              |||||||||
Db        538 GMGGMGGMD 546


RESULT 9
S39765
chaperonin 60 - Coxiella burnetii
C;Species: Coxiella burnetii
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 28-May-1999
C;Accession: S39765
R;Vodkin, M.H.; Williams, J.C.
J. Bacteriol. 170, 1227-1234, 1988
A;Title: A heat shock operon in Coxiella burnetii produces a major antigen homologous to
A;Reference number: S39764; MUID:88139182; PMID:3343219
A;Accession: S39765
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-552 <VOD>
A;Cross-references: GB:M20482; NID:g144996; PIDN:AAA23309.1; PID:g144998
C;Superfamily: chaperonin groEL

  Query Match             68.1%;  Score 1843.5;  DB 2;  Length 552;
  Best Local Similarity   66.5%;  Pred. No. 2.7e-82;
  Matches  365;  Conservative   82;  Mismatches   99;  Indels    3;  Gaps    3;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:| : |  :    :||||: |||||||||||||||::||||:| ||||||||||||
Db          1 MAAKVLKFSHEVLHAMSRGVEVLANAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| :||| |||||||||||||||  ||:| | || ||| :||||
Db         61 ELEDKFENMGAQMVKEVASRTSDDAGDGTTTATVLAQAILVEGIKAVIAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |   ||  |:|| :||| |||||||||||| :||:|||||| |||| ||||||:
Db        121 DKAVTAAVAELKKISKPCKDQKAIAQVGTISANSDKSIGDIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  |:|||||:|||||||||||: | : |  ||:||:||  :|||:::::::|:|| |
Db        181 SGLENALEVVEGMQFDRGYLSPYFINNQQNMSAELENPFILLVDKKISNIRELIPLLENV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  ||||:||||:|||||||||| :|| ::| |||||||:||||||:|||:||||:
Db        241 AKSGRPLLVIAEDIEGEALATLVVNNIRGVVKVAAVKAPGFGDRRKAMLQDIAVLTGGKV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  ||  ||:|||||: |::|||:||:| : ||| ||:||| :|| :||||
Db        301 ISEEVGLSLEAASLDDLGSAKRVVVTKDDTTIIDGSGDAGDIKNRVEQIRKEIENSSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:|||||||| |||||| |||||| |||||| ||||||:|||||||||:||||| | :
Db        361 DKEKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479
              | :| || ::  ::|:  |: | ||::  || ||  | | : ::| :||   ||  :|:|
Db        421 RVLKSLDSVEVENEDQRVGVEIARRAMAYPLSQIVKNTGVQAAVVADKVLNHKDVNYGYN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGG-M 538
              ||:||| |:|: |::|| |||| |||||||:| |::|||| : | |: ||:  |||| |
Db        481 AATGEYGDMIEMGILDPTKVTRTALQNAASIAGLMITTECMVTEAPK-KKEESMPGGGDM 539

Qy        539 GGMGGMDGM 547
              |||||| ||
Db        540 GGMGGMGGM 548


RESULT 10
F95967
probable heat shock protein groEL [imported] - Sinorhizobium meliloti (strain 1021) magap
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: F95967
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernand
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endos
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: F95967
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49406.1; PID:g15140892; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
```

hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: groEL5; SMb21566
A;Genome: plasmid
C;Superfamily: chaperonin groEL

Query Match 68.0%; Score 1841; DB 2; Length 542;
Best Local Similarity 68.3%; Pred. No. 3.5e-82;
Matches 373; Conservative 71; Mismatches 96; Indels 6; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   |||:: |||| |||||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFQTDARERMLRGVDVLANAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||::||| :|:|:|||||||||||||||| ||| | ||:| ||| :||||
Db         61 ELEDKFENMGAQMLREVASRTNDLAGDGTTTATVLAQAIVREGAKAVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | || || ||| :  :       |||||||||| || ||  :|||| ||| |||||||||
Db        121 DLAVDAVVKELKNNARKISKNSEIAQVGTISANGDTEIGRYLAEAMEKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:||||||||||:|| :||  ||::|||| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFITNQDKMRVELEDPYILIHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  :|||||||||||||||||||||||| |:| ||||||||:|||||||||||||||
Db        241 VQSGKPLLIIAEDVEGEALATLVVNKLRGGLKVAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              : || |||||:|:|  || ||:| |:|||||:|||   ||: |  |||||||||:|||
Db        301 VSEDLGIKLESVTLDMLGRAKKVSIEKENTTIIDGAGSKADIEGRTAQIRAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:|||||||:|||:||||||||||:|||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRVDDALHATRAAVEEGILPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479
              |::| || :| |::|:  |::::||::| |:|||| ||| ||||:| |:||  :  :|:|
Db        421 RAVKALDGLKTANNDQRVGVDLVRRAIEAPVRQIAENAGAEGSIIVGKLREKTEFSYGWN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539
              | : || ||   ||||| || | |||:||||| ||:|||  |||||: :    :| |
Db        481 AQTNEYGDLYAMGVIDPAKVVRTALQDAASVAGLLVTTEAMIAEKPKKEAAPALPAG--- 537

Qy        540 GMGGMD 545
                ||||
Db        538 --GGMD 541
```

RESULT 11
AD2660
60 KDA chaperonin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AD2660
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AD2660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41698.1; PID:g17739044; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: groEL
A;Map position: circular chromosome
C;Superfamily: chaperonin groEL

Query Match 68.0%; Score 1840.5; DB 2; Length 544;
Best Local Similarity 69.2%; Pred. No. 3.7e-82;
Matches 376; Conservative 69; Mismatches 95; Indels 3; Gaps 3;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   |||: :||| ||:|||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFGRSAREKMLKGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||:|:||| ||:|||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|     |     :|:||||||||| :  ||  |||| :||  ||||||||
Db        121 DLAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:||||:|||||||||||||||||| :|:: ||| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  :|||:|||||||||||||||||||| |:: ||||||||:|||||||||||||||
Db        241 VQTGKPLVIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || ||||:|:|  || :|:|  | ||||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGKSKKVSISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:|||||::|||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479
              ||   :  :|  :||: ||:|||:|::|:  :|||| ||| | |||| |: :  :| :|:|
Db        421 RSSTKI-TVKGVNDDQEAGINIVRKALQSLVRQIAENAGDEASIVVGKILDKNEDNYGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538
              | :||| |||  |::|| || | ||||||||||:|||  ||| |: :  || ||||||
Db        480 AQTGEYGDLIALGIVDPVKVVRTALQNAASVASLLITTEAMIAELPKKESAMPQMPGGGM 539

Qy        539 GGM 541
              |||
Db        540 GGM 542
```

RESULT 12
B97442
60K chaperonin (protein cpn60) (groEL protein) [imported] - Agrobacterium tumefaciens (s
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: B97442
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B97442
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86491.1; PID:g15155641; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1220
A;Map position: circular chromosome
C;Superfamily: chaperonin groEL

Query Match 68.0%; Score 1840.5; DB 2; Length 544;

Best Local Similarity 69.2%; Pred. No. 3.7e-82;
Matches 376; Conservative 69; Mismatches 95; Indels 3; Gaps 3;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||||: :||| ||:||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFGRSAREKMLKGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||:|:||| ||:|||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|     |      :|:|||||||| :  ||  ||||| :||  ||||||||
Db        121 DLAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||  |:|||||:|||||||||||||||||| :|:: ||| :|||:::::  |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  :||:|||||||||||||||||||| |:: |||||||:|||||||||||||||||
Db        241 VQTGKPLVIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || :|:| | ||||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGKSKKVSISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:|||||::||||||||||:||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479
              ||   :  :|  :||: ||:||:|::|:  :|||| ||| | |||| |: :  :| :|:|
Db        421 RSSTKI-TVKGVNDDQEAGINIVRKALQSLVRQIAENAGDEASIVVGKILDKNEDNYGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538
              | :||| |||  |::|| || | ||||||||||:|||  ||| |: :  || ||||||
Db        480 AQTGEYGDLIALGIVDPVKVVRTALQNAASVASLLITTEAMIAELPKKESAMPQMPGGGM 539

Qy        539 GGM 541
              |||
Db        540 GGM 542
```

RESULT 13
S23918
groEL protein - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C;Accession: S23918
R;Segal, G.; Ron, E.Z.
submitted to the EMBL Data Library, August 1992
A;Description: Cloning and sequencing of the GroE operon of Agrobacterium tumefaciens C5
A;Reference number: S23917
A;Accession: S23918
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <SEG>
A;Cross-references: EMBL:X68263; NID:g1019913; PIDN:CAA48331.1; PID:g39095
C;Superfamily: chaperonin groEL

Query Match 67.5%; Score 1828.5; DB 2; Length 544;
Best Local Similarity 68.9%; Pred. No. 1.4e-81;
Matches 374; Conservative 70; Mismatches 96; Indels 3; Gaps 3;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||:   | ||||: :||| ||:|||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKPGASAREKMLKGVDILADAVKVTLGPKGRNVVIDKSFGAPPITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||:|:||| ||:|||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|     |      :|:|||||||| :  ||  ||||| :||  ||||||||
Db        121 DLAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||  |:|||||:|||||||||||||||||| :|:: ||| :|||:::::  |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  :||:|||||||||||||||||||| |:: |||||||:|||||||||||||||||
Db        241 VQTGKPLVIIAEDVEGERLATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || :|:| | ||||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGKSKKVSISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:|||||::||||||||||:||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREP-KDGFGFN 479
              ||   :  :|  :||: ||:||:|::|:  :|||| ||| | |||| |: :  :| :|:|
Db        421 RSSTKI-TVKGVNDDQEAGINIVRKALQSLVRQIAENAGDEASIVVGKILDKNEDNYGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGM 538
              | :||| |||  |::|| || | ||||||||||:|||  ||| |: :  || ||||||
Db        480 AQTGEYGDLIALGIVDPVKVVRTALQNAASVASLLITTEAMIAELPKKESAMPQMPGGGM 539

Qy        539 GGM 541
              |||
Db        540 GGM 542
```

RESULT 14
B83098
GroEL protein PA4385 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83098
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83098
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-547 <STO>
A;Cross-references: GB:AE004854; GB:AE004091; NID:g9950606; PIDN:AAG07773.1; GSPDB:GN0013
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: groEL; PA4385
C;Superfamily: chaperonin groEL

Query Match 67.4%; Score 1825.5; DB 2; Length 547;
Best Local Similarity 66.8%; Pred. No. 2e-81;
Matches 367; Conservative 75; Mismatches 102; Indels 5; Gaps 2;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||:|:  ||: ||:||| ||||||||||::||||:| ||||||||||||
Db          1 MAAKEVKFGDSARKKMLVGVNVLADAVKATLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||:||:|| | :| ||||||||||||||||  ||:| |||| ||| :||||
Db         61 ELKDKFENMGAQLVKDVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||| ||:  :| :: ||  | | ||||||||||| :|| |||||| |||| |||||||
Db        121 DKATVAIVAQLKELAKPCADTKAIAQVGTISANSDESIGQIIAEAMEKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  | |||||:||||||||||  |: |  |||:|  :|  :|||:::::|||:|| |
```

```
Db        181 SGLENELSVVEGMQFDRGYLSPYFVNKPDTMAAELDSPLLLLVDKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 AKAGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEEVGLSLEGATLEHLGNAKRVVINKENTTIIDGAGVQADIEARVLQIRKQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Db        421 RALQAIEGLKGDNEEQNVGIALLRRAVESPLRQIVANAGDEPSVVVDKVKQGSGNYGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGG--GM 538

Db        481 ATGVYGDMIEMGILDPAKVTRSALQAAASIGGLMITTEAMVAEIVE---DKPAMGGMPDM 537

Qy        539 GGMGGMDGM 547

Db        538 GGMGGMGGM 546
```

```
RESULT 15
S22347
groEL - Brucella abortus
C;Species: Brucella abortus
C;Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 26-Aug-1999
C;Accession: S22347
R;Gor, D.; Mayfield, J.E.
Biochim. Biophys. Acta 1130, 120-122, 1992
A;Title: Cloning and nucleotide sequence of the Brucella abortus groE operon.
A;Reference number: S22346; MUID:92182006; PMID:1347461
A;Accession: S22347
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-546 <GOR>
A;Cross-references: EMBL:M82975; NID:g144109; PIDN:AAA22997.1; PID:g144111
C;Superfamily: chaperonin groEL

  Query Match             67.3%;  Score 1823;  DB 2;  Length 546;
  Best Local Similarity   68.9%;  Pred. No. 2.6e-81;
  Matches  378;  Conservative   64;  Mismatches   99;  Indels    8;  Gaps    5;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKDVKFGRTAREKMLRGVDILADAVKVTLGPKGRNVVIEKSFGAPRITKDGVSVAKEV 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELEDKFENMGAQMLREVASKTNDTAGDGTTTATVLGQAIVQEGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DLAVNEVVAELLKKAKKINTSEEVAQVGTISANAEAEIGKMIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQALLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Db        241 VQTSKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Db        301 ISEDLGIKLESVTLDMLGRAKKVSISKENTTIVDGAGQKAEIDARVGQIKQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGTALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVRE-PKDGFGF 478

Db        421 RASTKI--TAKGVNADQEAGINIVRRAIQAPARQITTNAGEEASVIVGKILENTSETFGY 478

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--G 536

Db        479 NTANGEYGDLISLGIVDPVKVVRTALQNAASVAGLLITTEAMIAEL--PKKDAAPAGMPG 536

Qy        537 GMGGMGGMD 545

Db        537 GMGGMGGMD 545

Search completed: January 28, 2004, 13:17:49
Job time : 22 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:10:35 ; Search time 17 Seconds
(without alignments)
1515.920 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1891.5	69.8	547	1	CH60_RALSO	Q8y1p8 ralstonia s
2	1880	69.4	546	1	CH60_ACEAC	Q8gbd2 acetobacter
3	1872.5	69.1	546	1	CH60_BURCE	Q9zfe0 burkholderi
4	1868.5	69.0	546	1	CH60_BURPS	Q9f712 burkholderi
5	1868.5	69.0	546	1	CH60_BURVI	Q9zfd8 burkholderi
6	1868	69.0	546	1	CH60_RHOPA	Q93mh1 rhodopseudo
7	1867.5	69.0	546	1	CH60_BURTH	P58723 burkholderi
8	1867.5	69.0	547	1	CH60_BARHE	O33963 bartonella
9	1867	68.9	545	1	CH61_RHIME	P35469 rhizobium m
10	1866	68.9	551	1	CH60_AMOPS	P26004 amoeba prot
11	1864.5	68.9	547	1	CH60_BARQU	O33964 bartonella
12	1864	68.8	545	1	CH64_RHIME	Q92zq4 rhizobium m
13	1862.5	68.8	547	1	CH60_XYLFA	Q9pfp2 xylella fas
14	1852.5	68.4	540	1	CH60_RHOMR	Q9xca9 rhodothermu
15	1852.5	68.4	546	1	CH60_CHRVI	P31293 chromatium
16	1847	68.2	546	1	CH60_XANAC	Q8ppz1 xanthomonas
17	1845	68.1	545	1	CH60_RHOCA	P95678 rhodobacter
18	1844.5	68.1	546	1	CH60_RHILV	P34939 rhizobium l
19	1844	68.1	546	1	CH60_XANCH	Q8rit7 xanthomonas
20	1843.5	68.1	547	1	CH60_BORPE	P48210 bordetella
21	1843.5	68.1	552	1	CH60_COXBU	P19421 coxiella bu
22	1842	68.0	546	1	CH60_XANCP	Q8pd23 xanthomonas
23	1841	68.0	542	1	CH65_RHIME	P35471 rhizobium m
24	1840.5	68.0	544	1	CH60_AGRT5	P30779 agrobacteri
25	1834.5	67.7	546	1	CH61_RHOSH	P20110 rhodobacter
26	1834.5	67.7	552	1	CH60_PSEST	O33500 pseudomonas
27	1834	67.7	542	1	CH62_RHILE	Q91691 rhizobium l
28	1833.5	67.7	539	1	CH61_BRAJA	P77829 bradyrhizob
29	1832.5	67.7	545	1	CH60_PARDE	Q9z462 paracoccus
30	1828.5	67.5	545	1	CH60_CHLTE	Q8kf02 chlorobium
31	1826	67.4	542	1	CH62_RHILO	Q98ih9 rhizobium l
32	1825.5	67.4	547	1	CH60_PSEAE	P30718 pseudomonas
33	1823	67.3	546	1	CH60_BRUAB	P25967 brucella ab
34	1819	67.2	546	1	CH60_BRUSU	Q8fx87 brucella su
35	1818	67.1	549	1	CH62_BRAJA	P35861 bradyrhizob
36	1816	67.1	542	1	CH62_RHIME	P35470 rhizobium m
37	1814.5	67.0	544	1	CH60_NEIMA	P57006 neisseria m
38	1814	67.0	546	1	CH60_BRUME	Q8yb53 brucella me
39	1812.5	66.9	551	1	CH64_RHILO	Q983s4 rhizobium l
40	1810	66.8	552	1	CH63_RHILO	Q98ax9 rhizobium l
41	1808.5	66.8	544	1	CH60_NEIMB	P42385 neisseria m
42	1805.5	66.7	543	1	CH60_BARBA	P35635 bartonella
43	1804.5	66.6	547	1	CH60_ALTHA	Q9xau7 alteromonas
44	1803.5	66.6	549	1	CH65_RHILO	Q981j9 rhizobium l
45	1802.5	66.6	544	1	CH60_NEIGO	P29842 neisseria g

ALIGNMENTS

RESULT 1
CH60_RALSO
ID CH60_RALSO STANDARD; PRT; 547 AA.
AC Q8Y1P8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL OR MOPA OR RSC0642 OR RS01546.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC ----------

CC ----------
DR EMBL; AL646060; CAD14172.1; -.
DR HAMAP; MF_00600; -; 1.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding; Complete proteome.
SQ SEQUENCE 547 AA; 57404 MW; 0487E4650867C25A CRC64;

Query Match 69.8%; Score 1891.5; DB 1; Length 547;
Best Local Similarity 69.5%; Pred. No. 2.1e-83;
Matches 380; Conservative 70; Mismatches 96; Indels 1; Gaps 1;

```
Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:::|   || |:  ||: ||||||||||||||||||:|:||| | :||||||||||
Db          1 MAAKDVVFGDAARAKMVEGVNILANAVKVTLGPKGRNVVLERSFGGPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|| :|||||||||| |||| ||||||||||||:| |||:| |||| ||| :||||
Db         61 ELKDKLQNMGAQMVKEVASKTSDNAGDGTTTATVLAQSIVREGMKYVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  |:|||   ||||||| |||||| :||  |||||| |||| ||||||:
Db        121 DKAVAAAVEELKKISKPTTTSKEIAQVGAISANSDESIGARIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | ||  |||||||:|||||||||: |||| | :||||::|  :|||::::|:|:||||
Db        181 KSLEDELDVVEGMQFDRGYLSPYFINNPEKQVVQLDNPFVLLFDKKISNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||:|||||||||||||| :|| |:  |||||||:|||||||||||||||:
Db        241 AKAGRPLLIVAEDVEGEALATLVVNNIRGILKTAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: || |||| | ||||||:|||| : :|:|||||:|||||| :|||
Db        301 IAEEVGLTLEKATLNDLGQAKRVEIGKENTTIIDGAGDARNIEARVKQVRAQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||||||:||||||||||| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:  ::  :| |: |: ||: |: |::|||||||  ||| | |:||  |   |  :|:||
Db        421 RARALISGLKGANADQDAGIKIVLRAMEEPLRQIVTNAGDEASVVVANVIAGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              ::||| ||:: ||:|| |||| ||||||||||:|||:||:|| |:      || |||||
Db        481 STGEYGDLVEMGVLDPTKVTRTALQNAASVASLMLTTDCAVAELPKDDAAPAMP-GGMGG 539

Qy        541 MGGMDGM 547
              |||||||
Db        540 MGGMDGM 546
```

RESULT 2
CH60_ACEAC
ID CH60_ACEAC STANDARD; PRT; 546 AA.
AC Q8GBD2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL.
OS Acetobacter aceti.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acetobacter.
OX NCBI_TaxID=435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 3283;
RA Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
RA Yanagida F.;
RT "Cloning and Characterization of groEL Operon in Acetobacter aceti.";
RL J. Biosci. Bioeng. 94:140-147(2002).
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC   proper assembly of unfolded polypeptides generated under stress
CC   conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC   7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB081586; BAC16232.1; -.
DR HAMAP; MF_00600; -; 1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 58106 MW; 3E476AF42F2CD5CC CRC64;

```
  Query Match             69.4%;  Score 1880;  DB 1;  Length 546;
  Best Local Similarity   69.8%;  Pred. No. 7.2e-83;
  Matches  381;  Conservative   69;  Mismatches   94;  Indels    2;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: | | ||:::  ||| ||:|||||||||||||::||||:| ||||||||||||
Db          1 MAAKDVKFGADARQRMRAGVDILADAVKVTLGPKGRNVVLDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||::||| ||:||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMLREVASKTNDIAGDGTTTATVLAQAIVREGHKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| | :||  |  | ||||||||| :: || :|:||| ||| ||||||||
Db        121 DKAVAVVIEELKKNAKKVTTPAETAQVGTISANGESEIGQMISEAMQKVGSEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  :| |||||||:|||||:|||||||||||  :|:|||||  :|||::|:: |||:|| |
Db        181 KHFQTELDVVEGMQFDRGYISPYFVTNPEKMTADLENPYILIHEKKLSSLQPMLPLLESV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  ||||||||||:|||||||||||||| |:: ||||||||:||  ||||||||||:
Db        241 VQSGRPLLIIAEDVDGEALATLVVNKLRGGLKIAAVKAPGFGDRRNVMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || ||||| |:|: ||||||:| |||||||||||||::||| ||||||||||||||||
Db        301 ISEDLGIKLETVTLNMLGTAKKVHIDKENTTIVDGAGKADDIKGRVKQIRAQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:||:||||:|||:|||||||||||||||||||
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKERKDRVDDALHATRAAVEEGIVPGGGTALA 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479
              |:   |: :   :||:  | :||||:|: |||||| ||| :|:::  || |  | |||:
Db        421 RATLKLEGLHYHNDDQRVGGDIIRRALQAPLRQIAHNAGEDGAVIANKVLENSDYNFGFD 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539
              | :|||::|::||:||| || | |||:||||| ||:|||  :||:|| ||  |  |  ||
Db        481 AQAGEYKNLVEAGIIDPAKVVRTALQDAASVAGLLITTEAMVAERPE-KKAAPAGGPDMG 539

Qy        540 GMGGMD 545
              ||||||
Db        540 GMGGMD 545
```

RESULT 3
CH60_BURCE
ID CH60_BURCE STANDARD; PRT; 546 AA.
AC Q9ZFE0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL OR MOPA.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10744;
RA Zysk G., Splettstoesser W.D., Neubauer H.;

```
RT   "Nucleotide sequence comparison of the groE operon of Burkholderia
RT   spp.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF104907; AAC79087.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  56980 MW;  174B9934345E7315 CRC64;

  Query Match             69.1%;  Score 1872.5;  DB 1;  Length 546;
  Best Local Similarity   68.4%;  Pred. No. 1.6e-82;
  Matches  374;  Conservative   75;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:::|   || |:  ||: |||||||||||||||||:|:||| | :|||||||||||
Db          1 MAAKDVVFGDSARSKMVEGVNILANAVKVTLGPKGRNVVLERSFGGPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|| :|||||||||||| |||| |||||||||||||:| |||:| ||:| ||| :||||
Db         61 ELKDKLQNMGAQMVKEVASKTSDNAGDGTTTATVLAQSIVREGMKYVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  |:||    |||||||:||||||::||: ||||| |||| ||||||:
Db        121 DKAVAAAVEELKKISKPCTTNKEIAQVGSISANSDSSIGDRIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | |   |||||||:|||||||||: ||:| |  ||||::| ::||:::::|:||:||||
Db        181 KSLADELDVVEGMQFDRGYLSPYFINNPDKQVAVLDNPFVLLHDKKVSNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||||||||||||||||| :|| |: |||||||||:||||||||||||||:
Db        241 AKAGRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: || ||||: : ||||||:||||::  |:|||||:|||||| :|||
Db        301 IAEETGLTLEKATLAELGQAKRIEVGKENTTIIDGAGEAASIEARVKQVRAQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||:||| |||||| |||||| |||||:||||||||||| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:   :  :  |: |: ||: |: |::|||||||  | | | |:||  |    | :|:||
Db        421 RARTAIAGLTGANADQNAGIKIVLRAMEEPLRQIVTNGGEEASVVVAAVAAGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:||| |:::|||:|| |||| |||||||| |||||: |:||   ||:| |||||  ||
Db        481 ATGEYVDMVEAGVVDPTKVTRTALQNAASVAGLLLTTDAAVAEL--PKEDAPMPGGMPGG 538

Qy        541 MGGMDGM 547
              |||| ||
```

```
Db        539 MGGM-GM 544


RESULT 4
CH60_BURPS
ID   CH60_BURPS     STANDARD;      PRT;   546 AA.
AC   Q9F712;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=28450;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Woo P.C.Y., Leung P.K.L.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF287633; AAG32927.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57145 MW;  E9E9366EE8BD6ABD CRC64;

  Query Match             69.0%;  Score 1868.5;  DB 1;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 2.5e-82;
  Matches  373;  Conservative   75;  Mismatches   96;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:::|   || |:  ||: |||||||||||||||||:|:||| | :|||||||||||
Db          1 MAAKDVVFGDSARAKMVEGVNILANAVKVTLGPKGRNVVLERSFGGPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|| :|||||||||||| |||| |||||||||||||:| |||:| ||:| ||| :||||
Db         61 ELKDKLQNMGAQMVKEVASKTSDNAGDGTTTATVLAQSIVREGMKYVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  |:||    |||||||:||||||::||: ||||| |||| ||||||:
Db        121 DKAVAAAVEELKKISKPCTTNKEIAQVGAISANSDSSIGDRIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | |   |||||||:|||||||||: ||:| |  ||||::| ::||:::::|:||:||||
Db        181 KSLADELDVVEGMQFDRGYLSPYFINNPDKQVAVLENPFVLLHDKKVSNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||||||||||||||||| :|| |: |||||||||:||||||||||||||:
Db        241 AKAGRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
```

```
                | |: |: ||  :|: || |||: : ||||||:||||:: :|:|||||||  |||| :|||
Db        301 IAEETGLTLEKATLAELGQAKRIEVGKENTTIIDGAGEAVNIEARVKQIRTQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||||||:|||||||||||||||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:   :  :   : |: ||: |: |::|||||||  | | | |:||  |   |  :|:||
Db        421 RARTAIASLTGVNADQNAGIKIVLRAMEEPLRQIVTNGGEEASVVVAAVAAGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:||| |:::|||:|| |||| ||||||||| |||||: |:||   ||:| |||||  ||
Db        481 ATGEYVDMVEAGVVDPTKVTRTALQNAASVAGLLLTTDAAVAEL--PKEDAPMPGGMPGG 538

Qy        541 MGGMDGM 547
              |||| ||
Db        539 MGGM-GM 544


RESULT 5
CH60_BURVI
ID   CH60_BURVI     STANDARD;      PRT;   546 AA.
AC   Q9ZFD8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Burkholderia vietnamiensis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=60552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 11319;
RA   Zysk G., Splettstoesser W.D., Neubauer H.;
RT   "Nucleotide sequence comparison of the groE operon of Burkholderia
RT   spp.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF104908; AAC79089.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57010 MW;  47FA893959221218 CRC64;

  Query Match             69.0%;  Score 1868.5;  DB 1;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 2.5e-82;
  Matches  373;  Conservative   75;  Mismatches   96;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:::|   || |:  ||: ||||||||||||||||:|:||| | :|||||||||||
Db          1 MAAKDVVFGDSARSKMVEGVNILANAVKVTLGPKGRNVVLERSFGGPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|| :|||||||||| |||| |||||||||||||:| |||:| ||:| ||| :||||
Db         61 ELKDKLQNMGAQMVKEVASKTSDNAGDGTTTATVLAQSIVREGMKYVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :||  |:||    |||||||:||||||::||: ||||| |||| ||||||:
Db        121 DKAVAAAVEELKKISKPCTTNKEIAQVGSISANSDSSIGDRIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | |   |||||||:|||||||||: ||:| |  ||||::| ::||:::::|:||:||||
Db        181 KSLADELDVVEGMQFDRGYLSPYFINNPDKQVAVLDNPFVLLHDKKVSNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  ||||||||||||||||||| :|| |: |||||||||:||||||||||||||||:
Db        241 AKAGRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: || |||: : ||||||:||||::  |:|||||:| |||| :|||
Db        301 IAEETGLTLEKATLAELGQAKRIEVGKENTTIIDGAGEAASIEARVKQVRTQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||||||:|||||||||||||||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:   :  :  |: |: ||: |: |::|||||||  | | | |:||  |   |  :|:||
Db        421 RARTAIAGLTGANADQNAGIKIVLRAMEEPLRQIVTNGGEEASVVVAAVAAGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:||| |:::|||:|| |||| ||||||||| |||||: |:||   ||:| |||||  ||
Db        481 ATGEYVDMVEAGVVDPTKVTRTALQNAASVAGLLLTTDAAVAEL--PKEDAPMPGGMPGG 538

Qy        541 MGGMDGM 547
              |||| ||
Db        539 MGGM-GM 544


RESULT 6
CH60_RHOPA
ID   CH60_RHOPA     STANDARD;      PRT;   546 AA.
AC   Q93MH1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Rhodopseudomonas palustris.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Rhodopseudomonas.
OX   NCBI_TaxID=1076;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xiao M., Zhu C.R., Qian X.M., Zheng P., Chen Y.Y.;
RT   "Cloning and Sequencing of the groESL Operon of Rhodopseudomonas
RT   palustris.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF406639; AAK94943.1; -.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57937 MW;  163C7115D68FFFF8 CRC64;

  Query Match             69.0%;  Score 1868;  DB 1;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 2.7e-82;
  Matches  374;  Conservative   75;  Mismatches   95;  Indels    4;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: ||  ||::: |||: ||:||||||||||||||||:||||:| ||||||||||||
Db          1 MAAKDVKFDTDARDRMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||||||| :|:| |||||||||||||| :||:| |||| ||| :||||
Db         61 ELSDKFENMGAQMVKEVASRTNDEAGDGTTTATVLAQAIIKEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | |   | : :   ::|  || |:||||||||| :  ||  ||:|| |||  |||||||
Db        121 DLATAKVVESIKAASRPVNDQHEVAQVGTISANGEAQIGRFIADAMQKVGNEGVITVEEN 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||||| ::|||||:|||||||||||||| :||  ||:: :|| :|||::|:: |:|:|| |
Db        181 KGLETEVEVVEGMQFDRGYLSPYFVTNADKMTAELEDVFILLHEKKLSSLQPMVPLLESV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  ||||:|||||||||||||||||||| |:: |||||||:|||||:|||||||:
Db        241 IQAQRPLLIVAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLQDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |:|||||::  || ||:|  |:|:|||||||||:  :|:||| ||| |||||:|||
Db        301 ISEDLGMKLENVTIDMLGRAKKVSINKDNTTIVDGAGEKAEIEARVSQIRTQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||:||| |||||| ||  || |:||:||||:|||||||||:|||| ||| | :
Db        361 DREKLQERVAKLAGGVAVIRVGGMTEVEVKERKDRVDDALNATRAAVQEGIVVGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479
              :: |||| :   : |: ||: |:||:|| |||||| ||| :||:|  ||||  | ||||
Db        421 QAGKVLDGLTGENPDQNAGITIVRRALEAPLRQIAQNAGVDGSVVAGKVRESDDKAFGFN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539
              | : || |: | ||||| || | ||::|||||||:|||  ||:|||||      | ||||
Db        481 AQTEEYGDMFKFGVIDPAKVVRTALEDAASVASLLITTEAMIADKPEPKS---APAGGMG 537

Qy        540 GMGGMDGM 547
              ||||||||
Db        538 GMGGMDGM 545


RESULT 7
CH60_BURTH
ID   CH60_BURTH     STANDARD;      PRT;   546 AA.
AC   P58723;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Burkholderia thailandensis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Burkholderiaceae; Burkholderia.
OX   NCBI_TaxID=57975;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Woo P.C.Y., Woo G.K.S., Yuen K.Y.;
RT   "Burkholderia thailandensis chaperonin GroEL gene, complete
RT   sequence.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF454383; AAL49762.1; -.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57145 MW;  913CD01A5C131A06 CRC64;

  Query Match             69.0%;  Score 1867.5;  DB 1;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 2.8e-82;
  Matches  373;  Conservative   76;  Mismatches   95;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:::|   || |:  ||: |||||||||||||||||:|:||| | :||||||||||
Db          1 MAAKDVVFGDSARAKMVEGVNILANAVKVTLGPKGRNVVLERSFGGPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|| :|||||||||| |||| |||||||||||||:| |||:| ||:| ||| :||||
Db         61 ELKDKLQNMGAQMVKEVASKTSDNAGDGTTTATVLAQSIVREGMKYVASGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| |  :|| |:||     |||||||:||||||::||: ||||| |||| ||||||:
Db        121 DKAVAAAVEELKKISKPCTTNKEIAQVGSISANSDSSIGDRIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              | |   |||||||:|||||||||: ||:| |  |:||::| ::||:::::|:|:||||
Db        181 KSLADELDVVEGMQFDRGYLSPYFINNPDKQVAVLENPFVLLHDKKVSNIRDLLPVLEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||||||||||||||||| :|| |: |||||||||:|||||||||||||||:
Db        241 AKAGRPLLIIAEDVEGEALATLVVNNIRGILKTVAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: || ||||: : ||||||:||||:: :|:|||||| |||| :|||
Db        301 IAEETGLTLEKATLAELGQAKRIEVGKENTTIIDGAGEAVNIEARVKQIRTQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||:||| |||||| |||||| |||||| |||||:|||||||||||||||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGIVPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:   :  :   : |: ||: |: |::|||||||  | | | |:||  |   |  :|:||
Db        421 RARTAIAALTGVNADQNAGIKIVLRAMEEPLRQIVTNGGEEASVVVAAVAAGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |:||| |:::|||:|| |||| ||||||||| ||||: |:||   ||:| ||||| ||
Db        481 ATGEYVDMVEAGVVDPTKVTRTALQNAASVAGLLLTTDAAVAEL--PKEDAPMPGGMPGG 538

Qy        541 MGGMDGM 547
              |||| ||
```

```
Db        539 MGGM-GM 544


RESULT 8
CH60_BARHE
ID   CH60_BARHE     STANDARD;      PRT;   547 AA.
AC   O33963; O87267;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE   60).
GN   GROL OR GROEL OR MOPA.
OS   Bartonella henselae (Rochalimaea henselae).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bartonellaceae; Bartonella.
OX   NCBI_TaxID=38323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RX   MEDLINE=97419519; PubMed=9274034;
RA   Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;
RT   "Heat shock response and groEL sequence of Bartonella henselae and
RT   Bartonella quintana.";
RL   Microbiology 143:2807-2815(1997).
RN   [2]
RP   SEQUENCE OF 1-543 FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RA   Marston E.L., Sumner J.W., Regnery R.L.;
RT   "Evaluation of intraspecies genetic variation within the 60 kDa heat
RT   shock protein (groEL) gene of Bartonella species: a new phylogenetic
RT   analysis tool.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-408 FROM N.A.
RC   STRAIN=ATCC 49882 / Houston 1;
RX   MEDLINE=97373904; PubMed=9230387;
RA   Sumner J.W., Nicholson W.L., Massung R.F.;
RT   "PCR amplification and comparison of nucleotide sequences from the
RT   groESL heat shock operon of Ehrlichia species.";
RL   J. Clin. Microbiol. 35:2087-2092(1997).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78514; AAB69094.1; -.
DR   EMBL; AF014829; AAD04238.1; ALT_INIT.
DR   EMBL; U96734; AAB65637.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   547 AA;  57625 MW;  EDE25D566D5B10D2 CRC64;

  Query Match           69.0%;  Score 1867.5;  DB 1;  Length 547;
  Best Local Similarity 70.1%;  Pred. No. 2.8e-82;

  Matches  385;  Conservative   61;  Mismatches   96;  Indels    7;  Gaps    4;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |  :|||:| |||| |||||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFGREARERLLRGVDILANAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||::||| ||:|||||||||||||| ||| :|||| |||| ||| :||||
Db         61 ELEDKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQEGVKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  |   |     |  :   |||||||||||     || :||:|| |||  ||||||||
Db        121 DAAVDEVVANLFKKAKKIQTSAEIAQVGTISANGAAEIGKMIADAMEKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||||| |||| :||:|||| :|||:::::: :||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNAEKMVADLDDPYILIHEKKLSNLQSLLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  :|||||||||||||||||||||||| |:: ||||||||:|||||||||||||| |:
Db        241 VQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTSGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||||:|  || ||:| | ||||||:||||:  :| ||| ||: |||||:|||
Db        301 ISEDVGIKLENVTLDMLGRAKKVNISKENTTIIDGAGQKSEINARVNQIKVQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:|||||||:|||||||||||||||| |||||  :
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGTALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGFN 479
              |:   |  :| :: |: ||:||:||:|: | |||| ||| | :|:| || |   | ||:|
Db        421 RAANAL-TVKGSNPDQEAGINIVRRALQAPARQIATNAGEEAAIIVGKVLENNADTPGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDM---PMPGG 536
              |:||: |||  |::|| || | |||||||:||||:|||  :||  ||||  |||||
Db        480 TATGEFGDLIALGIVDPVKVVRSALQNAASIASLLITTEAMVAE--VPKKDTPVPPMPGG 537

Qy        537 GMGGMGGMD 545
              |||||||||
Db        538 GMGGMGGMD 546


RESULT 9
CH61_RHIME
ID   CH61_RHIME     STANDARD;      PRT;   545 AA.
AC   P35469;
DT   01-JUN-1994 (Rel. 29, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1).
GN   GROL1 OR GROEL1 OR GROEL-A OR R00792 OR SMC00913.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=93231539; PubMed=8097179;
RA   Rusanganwa E., Gupta R.S.;
RT   "Cloning and characterization of multiple groEL chaperonin-encoding
RT   genes in Rhizobium meliloti.";
RL   Gene 126:67-75(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Ogawa J.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
```

```
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- INDUCTION: By heat shock.
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94192; AAA26285.1; -.
DR   EMBL; U19726; AAA61955.1; -.
DR   EMBL; AL591785; CAC45364.1; -.
DR   PIR; JN0509; JN0509.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock; Multigene family;
KW   Complete proteome.
FT   CONFLICT     12      12       A -> G (IN REF. 1).
SQ   SEQUENCE   545 AA;  57687 MW;  4CEF358957E1B45A CRC64;

  Query Match             68.9%;  Score 1867;  DB 1;  Length 545;
  Best Local Similarity   70.7%;  Pred. No. 3e-82;
  Matches  388;  Conservative   58;  Mismatches   97;  Indels    6;  Gaps    4;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||||: |||| ||:|||||||||||||||:||||:| ||||||||||||
Db          1 MAAKEVKFGRSAREKMLRGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||:||| ||:||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|    |       |:||||||||| :  ||  ||||| |||  ||||||||
Db        121 DLAVAEVVKDLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  || |:|||||:|||||||||||||||||| :|:: :|| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAFILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  :|||||||||||||||||||||| |:: |||||||:|||||||||||||||
Db        241 VQTGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|   || ||:| | |||||||||:  ||: || ||:|||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSITKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:||||||::||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGF 478
              | |:|:   :|  :||: ||:||:||:|: | |||  ||| | |||| |: |   | ||:
Db        421 RSSVKI--TVKGENDDQDAGVNIVRRALQSPARQIVENAGDEASIVVGKILEKNTDDFGY 478

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGM 538
              || :||| |:|  |:||| || | |||:||||||||:|||  |||   |||| |   |||
Db        479 NAQTGEYGDMIAMGIIDPVKVVRTALQDAASVASLLITTEAMIAEL--PKKDAPAMPGGM 536

Qy        539 GGMGGMDGM 547
              ||||||| |
Db        537 GGMGGMDMM 545


RESULT 10
CH60_AMOPS
ID   CH60_AMOPS     STANDARD;      PRT;   551 AA.
AC   P26004;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Amoeba proteus symbiotic bacterium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Legionellaceae.
OX   NCBI_TaxID=2728;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ahn T.I., Leeu H.K., Kwak I.H., Jeon K.W.;
RT   "Nucleotide sequence and temperature-dependent expression of groEL
RT   gene isolated from symbiotic bacteria of Amoeba proteus.";
RL   Endocyt. Cell Res. 8:33-44(1991).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86549; AAC09381.1; -.
DR   PIR; JC2562; JC2562.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   551 AA;  58332 MW;  A25895CBF9585DA0 CRC64;

  Query Match             68.9%;  Score 1866;  DB 1;  Length 551;
  Best Local Similarity   67.9%;  Pred. No. 3.4e-82;
  Matches  373;  Conservative   74;  Mismatches   96;  Indels    6;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||: |   ||::: ||: ||: || |:|| |||||:|:|||:| :||||||||||||
Db          2 AKELRFGDDARQQMLAGVNALADRVKATMGPSGRNVVLERSFGAPTVTKDGVSVAKEIEF 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              |::|:|||||||||| |||| |||||||||||::|   || | |||| ||| :||||||
Db         62 ENRFKNMGAQMVKEVAAKTSDTAGDGTTTATVLARSIVVEGHKAVAAGMNPMDLKRGIDK 121
```

```
Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
Db        122 AVTAITKELQKMSKPCKDGKAIAQVGTISANSDQAIGSIIAEAMEKVGKEGVITVEDGNG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
Db        182 LENELSVVEGMQFDRGYISPYFINNQQNMSAELEHPFILLVDKKIATIRDMLSVLEAVAK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
Db        242 SGRPLLIVAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTNGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
Db        302 EEIGTSLETASLESLGTAKRIVVTKENTTIIDGEGKATEINARIAQIRAQMEETSSDYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
Db        362 EKLQERVAKLAGGVAVIKVGAATEIEMKEKKARVEDALHATRAAVEEGIVAGGGVALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
Db        422 QKVLDGLKGDNADQDMGINILRRAIESPLRQIVANAGYESSVIVNKVAEHKDNFGFNAAT 481

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG----GGM 538
Db        482 GQYGDMVEMGILDPTKVTRTALQNAASVRSLMLTTECMVADL--PKKDEGMAGAGDMGGM 539

Qy        539 GGMGGMDGM 547
Db        540 GGMGGMGGM 548
```

```
RESULT 11
CH60_BARQU
ID   CH60_BARQU     STANDARD;      PRT;   547 AA.
AC   O33964;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE   60).
GN   GROL OR GROEL OR MOPA.
OS   Bartonella quintana (Rochalimaea quintana).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bartonellaceae; Bartonella.
OX   NCBI_TaxID=803;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51694;
RX   MEDLINE=97419519; PubMed=9274034;
RA   Haake D.A., Summers T.A., McCoy A.M., Schwartzman W.;
RT   "Heat shock response and groEL sequence of Bartonella henselae and
RT   Bartonella quintana.";
RL   Microbiology 143:2807-2815(1997).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78515; AAB69095.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock.
SQ   SEQUENCE   547 AA;  57610 MW;  8620CFC50BF2D6F4 CRC64;
```

```
  Query Match             68.9%;  Score 1864.5;  DB 1;  Length 547;
  Best Local Similarity   70.1%;  Pred. No. 4e-82;
  Matches  385;  Conservative   61;  Mismatches   96;  Indels    7;  Gaps    4;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
Db          1 MAAKEVKFGREARERLLRGVDILANAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
Db         61 ELEDKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQEGVKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
Db        121 DAAVEEVVGNLFKKAKKIQTSAEIAQVGTISANGAAEIGKMIADAMEKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNADKMVADLDDPYILIHEKKLSNLQSLLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
Db        241 VQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTSGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
Db        301 ISEDVGIKLENVTLDMLGRAKKVNISKENTTIIDGAGKKAEINARVNQIKVQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGTALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGFN 479
Db        421 RAANAL-AIKGSNPDQEAGINIVRRALQAPARQIATNAGEEAAIIVGKVLENNADTFGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDM---PMPGG 536
Db        480 TATGQFGDLIALGIVDPVKVVRSALQNAASIASLLITTEAMVAE--VPKKDTPMPPMPGG 537

Qy        537 GMGGMGGMD 545
Db        538 GMGGMGGMD 546
```

```
RESULT 12
CH64_RHIME
ID   CH64_RHIME     STANDARD;      PRT;   545 AA.
AC   Q92ZQ4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin 4 (Protein Cpn60 4) (groEL protein 4).
GN   GROL4 OR GROEL4 OR RA0395 OR SMA0744.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymA (megaplasmid 1).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
```

```
RX   MEDLINE=21396509; PubMed=11481432;
RA   Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA   Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA   Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA   Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA   Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT   "Nucleotide sequence and predicted functions of the entire
RT   Sinorhizobium meliloti pSymA megaplasmid.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE007230; AAK65053.1; -.
DR   PIR; C95311; C95311.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock; Multigene family; Plasmid;
KW   Complete proteome.
SQ   SEQUENCE   545 AA;  57701 MW;  49BA61F23CE70431 CRC64;

  Query Match             68.8%;  Score 1864;  DB 1;  Length 545;
  Best Local Similarity   70.5%;  Pred. No. 4.2e-82;
  Matches  387;  Conservative   59;  Mismatches   97;  Indels    6;  Gaps    4;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||||: |||| ||:||||||||||||||||:||||:| ||||||:|||||
Db          1 MAAKEVKFGRSAREKMLRGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||:||| ||:||||||||||||||||| ||| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  | |:|    |      |:|||||||| :  ||  ||||| ||| |||||||||
Db        121 DLAVAEVVKDLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||  |:|||||:||||||||||||||||| :|:: :|| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAFILLHEKKLSNLQAMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
               :  :|||||||||||||||||||||||| |:: ||||||||:||||||||||||||
Db        241 VQTGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || |||||:|:|  || ||:| | |||||||||||:  ||: || ||:||||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSITKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:|||||::|||||||||:|||||||| | :
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRIDDALNATRAAVQEGIVPGGGVALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPK-DGFGF 478
              | |:|:   :|  :||: ||:||:||:|: | |||  ||| | |||| |: |   | ||:
Db        421 RSSVKI--TVKGENDDQDAGVNIVRRALQSPARQIVENAGDEASIVVGKILEKNTDDFGY 478

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGM 538
              || :||| |:|  |:||| || | |||:||||||:|||  |||   |||| |   |||
Db        479 NAQTGEYGDMIAMGIIDPVKVVRTALQDAASVASLLITTEAMIAEL--PKKDAPAMPGGM 536

Qy        539 GGMGGMDGM 547
              |||||| |
Db        537 GGMGGMDMM 545


RESULT 13
CH60_XYLFA
ID   CH60_XYLFA     STANDARD;      PRT;   547 AA.
AC   Q9PFP2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR XF0615.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003907; AAF83425.1; -.
DR   PIR; F82783; F82783.
```

```
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Complete proteome.
SQ   SEQUENCE   547 AA;  57757 MW;  30BE7F937CA7A9D8 CRC64;

  Query Match             68.8%;  Score 1862.5;  DB 1;  Length 547;
  Best Local Similarity   66.8%;  Pred. No. 4.9e-82;
  Matches  366;  Conservative   80;  Mismatches   97;  Indels    5;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|||:|  ||| ::  ||: |||||| |||||||:||::||||||:|||||||||||||
Db          1 MAAKEIIFSEKARSRMVHGVNLLANAVKATLGPKGRHVVLDKSFGSPIITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||:|||| ||:| |||||||||||||: ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMLKEVASKTNDHAGDGTTTATVLAQALIREGCKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||:|    ||  |:||| | | |||| ||||||| :||||||||| |||| ||||:||
Db        121 DKAVIAAVTELKKISKPTSDDKAIAQVATISANSDESIGNIIAEAMKKVGKEGVITIEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||  |||||||:||||| |||| | :  : |||||||| ::|||:|::|:| :|: |
Db        181 TTLENELDVVEGMQFDRGYSSPYFINNQQSQIVELDNPYILLHDKKISSVRDLLTVLDAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              || ::||||:||:|||||||||||| :|| ::| |||||||:|||||||:|:||||
Db        241 AKESKPLLIVAEEVEGEALATLVVNNIRGIIKVCAVKAPGFGDRRKAMLEDMAVLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  : | || ||:| : ||||||:|| | :: |  |||||: |||||:|||
Db        301 ISEEVGLSLEKATTSHLGKAKKVRVSKENTTIIDGIGDNDAINGRVKQIKTQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||:||| |||||| |||||| |||:||| |||||||||::|||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVDDALLATRAAVEEGVIPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:|  : ::| |::|:  |: |  |::| |||:| |||| | |::: ||:| || ||:||
Db        421 RAITAISNLKGANEDQTHGIQIALRAMEAPLREIVANAGEEPSVILNKVKEGKDNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMP---GGG 537
              |:||: |::  |::|| |||| |||||||:| |::|||  :||  |||| | |   |||
Db        481 ATGEFGDMVNLGILDPTKVTRSALQNAASIAGLMITTEAMVAE--APKKDEPTPPAAGGG 538

Qy        538 MGGMGGMD 545
              ||||||||
Db        539 MGGMGGMD 546


RESULT 14
CH60_RHOMR
ID   CH60_RHOMR     STANDARD;      PRT;   540 AA.
AC   Q9XCA9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL.
OS   Rhodothermus marinus (Rhodothermus obamensis).
OC   Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC   Crenotrichaceae; Rhodothermus.
OX   NCBI_TaxID=29549;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ITI 376;
RA   Thorolfsdottir E.T.T., Backman V.M., Blondal T.,
RA   Thorbjarnardottir S.H., Palsdottir A., Hauksdottir H.,
RA   Kristjansdottir S., Eggertsson G.;
RT   "Heat shock in Rhodothermus marinus: cloning and sequence analysis of
RT   the groESL, dnaK and dnaJ genes.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions (By similarity).
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF145252; AAD37976.1; -.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   540 AA;  57666 MW;  6867448B7BEF18EC CRC64;

  Query Match             68.4%;  Score 1852.5;  DB 1;  Length 540;
  Best Local Similarity   69.2%;  Pred. No. 1.5e-81;
  Matches  371;  Conservative   71;  Mismatches   93;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:|  |:|  ||   | |||||||:|||||||||||||:||| ||:| :|||||:|||||
Db          1 MAAKQITFNADARMALKRGVDKLADAVKVTLGPKGRNVIIEKKFGAPTVTKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||| ||:|||||||| ||||:|||||||||||||||    |:| | || ||| :||||
Db         61 ELEDKLENVGAQMVKEVASKTSDVAGDGTTTATVLAQAILTAGLKSVTAGANPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||  |  ||  :::  :|:  |||| ||||| |  || :||:|  |||| ||||||||
Db        121 DKAVEVVVAELRKMSQEVQDKNRIAQVATISANGDKAIGQLIADAFEKVGKDGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              || ||||:|||||:|||||||||||||||: |   |:: ||| ::|||::|||:|||||:|
Db        181 KGTETTLEVVEGMQFDRGYLSPYFVTNPDTMEAVLEDAYILIHDKKISAMKDLLPILEKV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  |||||||||||||||||||||||| |:| ||||||||:|||||||||||||||||
Db        241 VQTGRPLLIIAEDVEGEALATLVVNKLRGVLKVAAVKAPGFGDRRKAMLEDIAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |::| :||| :|  || |:|:::||:|||| | |    ||||  ||| |||||:|||
Db        301 ISEEKGYRLENATLDYLGQAERIIVDKDNTTIVGGKGDPAQIKARANQIRQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||: :|||| |||||| |||||:|||||||||||||||| |::
Db        361 DREKLQERLAKLAGGVAVLKIGAATEPEMKEKKARVEDALHATRAAVEEGIVPGGGVAYI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:|  || ::  ::|:  |: |::|:||||||||||||:||||||::|:|  |||||
Db        421 RAIAALDKVEVENEDQKIGVQIVQRALEEPLRQIAANAGWEGSIVVQRVKEGQGDFGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKK-DMPMPG 535
              : |: :|:: ||||| || | ||:|||||| ||||||  :|||||  :|   | ||
Db        481 QTEEFGNLLEQGVIDPTKVARTALENAASVAGLLLTTEAVVAEKPEKEKAAAPSPG 536
```

```
RESULT 15
CH60_CHRVI
ID   CH60_CHRVI     STANDARD;      PRT;   546 AA.
AC   P31293;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR GROEL OR MOPA.
OS   Chromatium vinosum.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales;
OC   Chromatiaceae; Allochromatium.
OX   NCBI_TaxID=1049;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93186721; PubMed=8444812;
RA   Ferreyra R., Soncini F., Viale A.M.;
RT   "Cloning, characterization, and functional expression in Escherichia
RT   coli of chaperonin (groESL) genes from the phototrophic sulfur
RT   bacterium Chromatium vinosum.";
RL   J. Bacteriol. 175:1514-1523(1993).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=99008942; PubMed=9790891;
RA   Dionisi H.M., Viale A.M.;
RT   "Purification and characterization of Chromatium vinosum GroEL and
RT   GroES proteins overexpressed in Escherichia coli cells lacking the
RT   endogenous groESL operon.";
RL   Protein Expr. Purif. 14:275-282(1998).
CC   -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC       proper assembly of unfolded polypeptides generated under stress
CC       conditions.
CC   -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC       7 subunits.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M99443; AAA23319.1; -.
DR   PIR; B47073; B47073.
DR   HSSP; P06139; 1GRL.
DR   HAMAP; MF_00600; -; 1.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   546 AA;  57541 MW;  3250141881C04DD6 CRC64;

  Query Match             68.4%;  Score 1852.5;  DB 1;  Length 546;
  Best Local Similarity   68.2%;  Pred. No. 1.5e-81;
  Matches  374;  Conservative   70;  Mismatches  101;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              |::|:: |   || ::  ||: |||||||||||||||:|||||:| :|||||||||||
Db          1 MSAKDVKFGGDARVRMMEGVNILANAVKVTLGPKGRNVVLEKSFGAPTVTKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||||||||| ||||||||||||||||||: |||:| |||| ||| :|||:
Db         61 ELKDKFENMGAQMVKEVASKTSDIAGDGTTTATVLAQAMVREGLKAVAAGMNPMDLKRGM 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              |||| | |:||  ::||       |||||||||||| :|| |||||| |||| ||||||:

Db        121 DKAVEAATEELKKLSKPCPRPMAIAQVGTISANSDDSIGTIIAEAMEKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
                |:  |||||||:||||||||||: | :    ||| ||||  :|||::::|:||:|| |
Db        181 TSLQNELDVVEGMQFDRGYLSPYFINNQQSQSAELDAPYILLYDKKISNIRDLLPVLEGV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||||||||| :|| ::| ||||||||:||||||:|||||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTIRGIVKVCAVKAPGFGDRRKAMLQDIAILTGATV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |: ||  :|: |||||||| : |: |||:||:|   |||||  :|||||:|||||||
Db        301 ISEEVGLSLEKATLTDLGTAKRVQVGKDETTIIDGSGSEIDIKARCEQIRAQVEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| |||||| |||||| ||||||:||||||||||||||| | |
Db        361 DREKLQERLAKLAGGVAVIKVGAATEIEMKEKKARVEDALHATRAAVEEGIVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |:|  : |:| |: |:  |: | ||::|||||| |||| | |::: || |    ||:||
Db        421 RAIAAVKDLKGANHDQDVGIAIARRAMEEPLRQIVANAGEEPSVILHKVAEGTGNFGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMP-MPGGGMG 539
              |:||| |::: |::|| |||| |||: ||| |::||| ||:  ||| | | |||||||
Db        481 ANGEYGDMVEMGILDPTKVTRSALQNSCSVAGLMITTEAMIAD--EPKDDAPAMPGGGMG 538

Qy        540 GMGGMDGM 547
                ||||  |
Db        539 DMGGMGMM 546


Search completed: January 28, 2004, 13:16:24
Job time : 19 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 28, 2004, 13:11:20 ; Search time 39 Seconds
(without alignments)
3625.970 Million cell updates/sec

Title: US-09-077-574A-2
Perfect score: 2708
Sequence: 1 MASKEILFDAKAREKLSRGV.....KDMPMPGGGMGGMGGMDGMY 548

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2708	100.0	548	2	O87888	O87888 lawsonia in
2	1880	69.4	546	2	Q8GBD2	Q8gbd2 acetobacter
3	1823	67.3	552	2	Q8KJ46	Q8kj46 rhizobium l
4	1819	67.2	546	16	Q8FX87	Q8fx87 brucella su
5	1804.5	66.6	545	2	Q8GB95	Q8gb95 heliobacill
6	1796.5	66.3	549	2	Q9AJB5	Q9ajb5 pseudoalter
7	1780.5	65.7	545	16	Q8CX48	Q8cx48 shewanella
8	1780.5	65.7	546	2	Q8GBB4	Q8gbb4 azotobacter
9	1768	65.3	539	2	Q9RC20	Q9rc20 bacillus sp
10	1768	65.3	546	2	O31198	O31198 leptospira
11	1754	64.8	539	2	Q9EZV4	Q9ezv4 bacillus st
12	1742	64.3	545	2	Q93FU8	Q93fu8 enterobacte
13	1736	64.1	538	2	O33688	O33688 primary end
14	1722.5	63.6	546	2	Q8RNU2	Q8rnu2 helicobacte
15	1714.5	63.3	548	2	Q8KIW5	Q8kiw5 buchnera sp
16	1711	63.2	543	2	Q8KZN2	Q8kzn2 thermus sp.
17	1710.5	63.2	548	2	Q8KIW7	Q8kiw7 buchnera sp
18	1706	63.0	555	2	Q9X603	Q9x603 primary end
19	1694	62.6	542	16	Q8CWW6	Q8cww6 streptococc
20	1691.5	62.5	536	2	Q8KIW2	Q8kiw2 buchnera sp
21	1691	62.4	537	2	Q8KJ14	Q8kj14 streptococc
22	1687	62.3	545	16	Q8CXL3	Q8cxl3 oceanobacil
23	1686	62.3	540	16	Q8CX22	Q8cx22 streptococc
24	1685	62.2	531	2	Q8GBD0	Q8gbd0 enterococcu
25	1685	62.2	540	16	Q8CX00	Q8cx00 streptococc
26	1683.5	62.2	537	2	Q8KIW4	Q8kiw4 buchnera sp
27	1681	62.1	536	2	Q8KJ10	Q8kj10 streptococc
28	1679.5	62.0	537	2	Q8KIW3	Q8kiw3 buchnera sp
29	1676.5	61.9	537	2	Q8KIW9	Q8kiw9 buchnera sp
30	1675	61.9	534	2	Q8GBB6	Q8gbb6 streptococc
31	1672.5	61.8	548	2	O08499	O08499 rhopalosiph
32	1671.5	61.7	548	2	O08500	O08500 sitobion av
33	1670	61.7	531	2	Q8GBC6	Q8gbc6 enterococcu
34	1669	61.6	531	2	Q8GBC4	Q8gbc4 enterococcu
35	1669	61.6	531	2	Q8GBC0	Q8gbc0 enterococcu
36	1668	61.6	534	2	Q8KIY3	Q8kiy3 streptococc
37	1667	61.6	534	2	Q8KJ12	Q8kj12 streptococc
38	1663.5	61.4	540	2	Q9X4R5	Q9x4r5 streptococc
39	1663	61.4	531	2	Q8GBC2	Q8gbc2 enterococcu
40	1663	61.4	535	2	Q8KJ08	Q8kj08 streptococc
41	1660	61.3	503	2	O87271	O87271 bartonella
42	1660	61.3	532	2	Q8GBC8	Q8gbc8 enterococcu
43	1659	61.3	532	2	Q8GBB8	Q8gbb8 enterococcu
44	1648.5	60.9	539	2	Q8GJ00	Q8gj00 fusobacteri
45	1644	60.7	532	16	Q8CWJ0	Q8cwj0 vibrio vuln

ALIGNMENTS

```
RESULT 1
O87888
ID   O87888      PRELIMINARY;      PRT;   548 AA.
AC   O87888;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   GroEL/HSP60 homolog (60 kDa chaperonin) (Protein Cpn60) (groEL
DE   protein).
OS   Lawsonia intracellularis.
OG   Plasmid pISI-2.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC   Desulfovibrionaceae; Lawsonia.
OX   NCBI_TaxID=29546;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98386497; PubMed=9720028;
RA   Dale C.J.H., Moses E.K., Ong C.C., Morrow C.J., Reed M.B., Hasse D.,
RA   Strugnell R.A.;
RT   "Identification and sequencing of the groE operon and flanking genes
RT   of Lawsonia intracellularis: use in phylogeny.";
RL   Microbiology 144:0-0(0).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; U45241; AAC36500.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone; Plasmid.
SQ   SEQUENCE   548 AA;  58605 MW;  6388C431E663E498 CRC64;
```

```
 Query Match           100.0%;  Score 2708;  DB 2;  Length 548;
 Best Local Similarity  100.0%;  Pred. No. 1.3e-124;
 Matches  548;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540

Qy        541 MGGMDGMY 548
              ||||||||
Db        541 MGGMDGMY 548
```

RESULT 2
Q8GBD2
ID Q8GBD2 PRELIMINARY; PRT; 546 AA.
AC Q8GBD2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE groEL protein.
GN GROEL.
OS Acetobacter aceti.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acetobacter.
OX NCBI_TaxID=435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO3283;
RA Okamoto-Kainuma A., Yan W., Kadono S., Tayama K., Koizumi Y.,
RA Yanagida F.;
RT "Cloning and Characterization of groEL Operon in Acetobacter aceti.";
RL J. Biosci. Bioeng. 94:140-147(2002).
DR EMBL; AB081586; BAC16232.1; -.
SQ SEQUENCE 546 AA; 58106 MW; 3E476AF42F2CD5CC CRC64;

```
  Query Match           69.4%;  Score 1880;  DB 2;  Length 546;
  Best Local Similarity  69.8%;  Pred. No. 3.5e-84;
  Matches  381;  Conservative   69;  Mismatches   94;  Indels    2;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: | | ||::: ||| ||:||||||||||||||::||||:| ||||||||||||
Db          1 MAAKDVKFGADARQRMRAGVDILADAVKVTLGPKGRNVVLDKSFGAPRITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              || |||||||||::||| ||:|||||||||||||||| ||| | |||| ||| :||||
Db         61 ELADKFENMGAQMLREVASKTNDIAGDGTTTATVLAQAIVREGHKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||  | :||    |       | |||||||||| :: || :|:||| ||| ||||||||
Db        121 DKAVAVVIEELKKNAKKVTTPAETAQVGTISANGESEIGQMISEAMQKVGSEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  :| |||||||:|||||:||||||||||  :|:||||| :|||::|:: |||:|| |
Db        181 KHFQTELDVVEGMQFDRGYISPYFVTNPEKMTADLENPYILIHEKKLSSLQPMLPLLESV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              :  ||||||||||:||||||||||||||| |:: ||||||||:||  ||||||||||:
Db        241 VQSGRPLLIIAEDVDGEALATLVVNKLRGGLKIAAVKAPGFGDRRNVMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | || ||||| |:|: |||||:| ||||||||||||||::||| ||||||||||||||||
Db        301 ISEDLGIKLETVTLNMLGTAKKVHIDKENTTIVDGAGKADDIKGRVKQIRAQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || :|| |:||:||||:|||:|||||||||||||||||||
Db        361 DREKLQERLAKLAGGVAVIRVGGSTEVEVKERKDRVDDALHATRAAVEEGIVPGGGTALA 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKD-GFGFN 479
              |:   |: :   :||:  | :||||:|: |||||| ||| :|::: || |  | |||:
Db        421 RATLKLEGLHYHNDDQRVGGDIIRRALQAPLRQIAHNAGEDGAVIANKVLENSDYNFGFD 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539
              | :|||::|::||:||| || | |||:||||| ||:|||  :|||:|| ||  |  |  ||
Db        481 AQAGEYKNLVEAGIIDPAKVVRTALQDAASVAGLLITTEAMVAERPE-KKAAPAGGPDMG 539

Qy        540 GMGGMD 545
              ||||||
Db        540 GMGGMD 545
```

RESULT 3
Q8KJ46
ID Q8KJ46 PRELIMINARY; PRT; 552 AA.
AC Q8KJ46;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable chaperonin groEL DF protein (60 kDa chaperonin) (Protein
DE Cpn60) (groEL protein).
GN GROEL.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R7A;
RX MEDLINE=21999272; PubMed=12003951;
RA Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,
RA Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,
RA Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;
RT "Comparative sequence analysis of the symbiosis island of
RT Mesorhizobium loti strain R7A.";
RL J. Bacteriol. 184:3086-3095(2002).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL; AL672115; CAD31231.1; -.
DR InterPro; IPR001844; Chaprnin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.

```
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   552 AA;  57747 MW;  7D0A8C53CD7FA6AB CRC64;

  Query Match           67.3%;  Score 1823;  DB 2;  Length 552;
  Best Local Similarity  67.8%;  Pred. No. 2.2e-81;
  Matches  374;  Conservative   73;  Mismatches   97;  Indels    8;  Gaps    4;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   |||:: |||: ||:|||||||||||||||:||||:| ||||||:|||||
Db          1 MAAKDVKFSRDARERMLRGVNILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVTVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||:||| ||:|||||||||||||||:| :|| | |||| ||| :||||
Db         61 ELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQSIVQEGHKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  |   |    |  :  :|:||||||: | | ::|  :|||||  |||  ||||||||
Db        121 DLAVTDVVATLIKNAKKIKTSEEVAQVGTIAGNGDESVGKMIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||  |:|||||:||||||||||||| :||| :|:: ||| :|||:::::: |||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNADKMVADLEDVYILLHEKKLSNLQTMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              : ::|||||:|||||||||||||||||| |:: ||||||||:|||||||||||||||:
Db        241 VQTSKPLLIISEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | ||  ||||||| |: || ||:|  | |||||||||||| |:|:  || ||: |||||:|||
Db        301 ISEDLGIKLENVGLNMLGRAKKVSISKENTTIVDGAGKKEEIQGRVAQIKQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:|||||||||| |||||| || ||| |:|||||||:|||||||||||||||||| | :
Db        361 DKEKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVPGGGVALL 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDG-FGFN 479
              |:   :: :  |: |: ||:||:||:|: | |||||||| | |||   |: | |   ||:|
Db        421 RASLSINAV-GANSDQAAGINIVRRALQAPARQIAANAGAEASIVAGKILENKGATFGYN 479

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKK----DMP--M 533
              | :||| |:|  |::|| || | |||:|||| ||:|||  ||| |: :     ||  |
Db        480 AQTGEYGDMIAMGIVDPVKVVRTALQDAASVAGLLVTTEAMIAEAPKKESAGGGGMPGGM 539

Qy        534 PGGGMGGMGGMD 545
              ||||||||||
Db        540 GGGGMGGMGGMD 551


RESULT 4
Q8FX87
ID   Q8FX87      PRELIMINARY;      PRT;   546 AA.
AC   Q8FX87;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Chaperonin, 60 kDa.
GN   GROEL OR BRA0195.
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR   EMBL; AE014521; AAN33401.1; -.
DR   TIGR; BRA0195; -.
KW   Complete proteome.
SQ   SEQUENCE   546 AA;  57501 MW;  2E4046EDC121A30A CRC64;

  Query Match           67.2%;  Score 1819;  DB 16;  Length 546;
  Best Local Similarity  68.7%;  Pred. No. 3.4e-81;
  Matches  377;  Conservative   64;  Mismatches  100;  Indels    8;  Gaps    5;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   ||||: |||| ||:|||||||||||||||:||||:| ||||||||||:
Db          1 MAAKDVKFGRTAREKMLRGVDILADAVKVTLGPKGRNVVIDKSFGAPRITKDGVSVAKEV 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||::||| ||:| |||||||||| ||| :|| | |||| ||| :||||
Db         61 ELEDKFENMGAQMLREVASKTNDTAGDGTTTATVLGQAIVQEGAKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              | ||  |  ||     :|:||||||||| :  || :|||| |||  ||||||||
Db        121 DLAVNEVVAELLKKAKKINTSEEVAQVGTISANGEAEIGKMIAEAMQKVGNEGVITVEEA 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |  ||  |:|||||:|||||||||||||| :|:: ||| :|||:::::: :||:|| |
Db        181 KTAETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYILLHEKKLSNLQALLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              : ::|||||||||||||||||||||| |:: ||||||||:|||||||||||||||:
Db        241 VQTSKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | ||  |||||:|:|  || ||:|  | ||||||||||:  :| ||| ||: ||||:|||
Db        301 ISEDLGIKLESVTLDMLGRAKKVSISKENTTIVDGAGQKAEIDARVGQIKQQIEETTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| || ||| |:|||||||:|||||||||||||| |||||:
Db        361 DREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGGTALL 420

Qy        421 R-SIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVRE-PKDGFGF 478
              | | |:    |  : |: ||:||:||::: | |||  ||| | |::| |: |    : ||:
Db        421 RASTKI--TAKGVNADQEAGINIVRRAIQAPARQITTNAGEEASVIVGKILENTSETFGY 478

Qy        479 NAASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--G 536
              | |:||| |||  |::|| || | |||||||| ||:|||  |||   ||||    |  |
Db        479 NTANGEYGDLISLGIVDPVKVVRTALQNAASVAGLLITTEAMIAEL--PKKDAAPAGMPG 536

Qy        537 GMGGMGGMD 545
              |||||||||
Db        537 GMGGMGGMD 545


RESULT 5
Q8GB95
ID   Q8GB95      PRELIMINARY;      PRT;   545 AA.
AC   Q8GB95;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   60 kDa chaperonin groEL (Fragment).
OS   Heliobacillus mobilis.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;
OC   Heliobacillus.
OX   NCBI_TaxID=28064;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22337798; PubMed=12446909;
RA   Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA   Blankenship R.E.;
RT   "Whole-genome analysis of photosynthetic prokaryotes.";
```

```
RL   Science 298:1616-1620(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA   Gerdes S., Kyrpides N., Overbeek R.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY142910; AAN87514.1; -.
FT   NON_TER     545    545
SQ   SEQUENCE   545 AA;  57927 MW;  B9F6DB02DBC18383 CRC64;

  Query Match             66.6%;  Score 1804.5;  DB 2;  Length 545;
  Best Local Similarity   65.9%;  Pred. No. 1.7e-80;
  Matches  360;  Conservative   84;  Mismatches   99;  Indels    3;  Gaps    3;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|  :|: :||  | :||: || ||:|||||||||||:|| ||||:|| |||::||||||
Db          2 AKMTVFNEEARRALEKGVNTLAEAVRVTLGPKGRNVVLEKKFGSPLITNDGVTIAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              |:  ||||||:||||| ||:|:|||||||||:||||| |||:| |||| ||| |||||:|
Db         62 ENPIENMGAQLVKEVATKTNDVAGDGTTTATILAQAIIREGMKNVAAGANPMVIKRGIEK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||    :|: :::||   :: |||| :||| :| |||::||||| |||| |||||||:||
Db        122 AVATCVEEIKNLSKPVESKEAIAQVASISA-ADATIGSLIAEAMEKVGKDGVITVEESKG 180

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              | |:||||| |||||:||| :|:|:||   |::|||| :|||:::||:||:||:: :
Db        181 FSTDLEVVEGMNFDRGYISPYMITDPDKMEAVLNDPYILITDKKISAIKDILPVLERIVQ 240

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              : |:||:||:||||||||||||||    | |||||||:||||||||||||||  :
Db        241 SGKQLMIISEDIEGEALATLVVNKLRGTFTCVGVKAPGFGDRRKAMLEDIAILTGGRVVS 300

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |:|||| ::  || |::| | || | |||||  :|||||: ||| | ||::|::|:
Db        301 EEVGLKLENATMDMLGRARQVRISKEETIIVDGAGSQDDIKARIAQIRRQYEESTSEFDK 360

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| |||||| |||||||:|||| |:|||||||||||||||||||||||| |
Db        361 EKLQERLAKLAGGVAVIQVGAATETELKEKKLRIEDALNATRAAVEEGIVPGGGTALVSI 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              | ||:::   ||  |: ||:|||||||||| |||||||:|||||:     | |||||:
Db        421 QKALDEVEVPAGDEATGVAIIKRSLEEPLRQIANNAGYEGSVVVEKVKNLPKGQGFNAAT 480

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              |||:| ||::|| |||| |||||||:|::|||||  :|:||| ||: |  ||| |||
Db        481 EVYEDMIAAGIVDPAKVTRSALQNAASIAAMLLTTEAIVADKPE-KKEAPGMPGGMPGMG 539

Qy        543 GMD-GM 547
              ||| ||
Db        540 GMDMGM 545


RESULT 6
Q9AJB5
ID   Q9AJB5      PRELIMINARY;      PRT;   549 AA.
AC   Q9AJB5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN   GROEL.
OS   Pseudoalteromonas sp. PS1M3.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Pseudoalteromonas.
OX   NCBI_TaxID=87791;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PS1M3;
RA   Kurusu Y., Nakamura T.;
RT   "Moleculer chaperone of Psychrophile.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AB057417; BAB39465.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   549 AA;  57346 MW;  7AFCC2E8A2F736AF CRC64;

  Query Match             66.3%;  Score 1796.5;  DB 2;  Length 549;
  Best Local Similarity   65.7%;  Pred. No. 4.3e-80;
  Matches  360;  Conservative   79;  Mismatches  108;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||:||   || |:  ||: |||||||||||||||::|||||||||||||||||||
Db          1 MAAKEVLFAGDARAKMLTGVNILANAVKVTLGPKGRNVVLDKSFGSPVITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||||||||||||||||| | :| |||||||||||||:|   ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQSIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKG-GVITVEE 179
              |||::|   ||  :: |  | | ||||||||||||  ||:|||||| |||:  |||||||
Db        121 DKAIIAAVAELKALSVPCADTKAIAQVGTISANSDKEIGDIIAEAMEKVGRNTGVITVEE 180

Qy        180 AKGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQ 239
              : ||  |||||||:|||||||||: | ||   |||||:||  :||:::::::|| ||
Db        181 GQSLENELDVVEGMQFDRGYLSPYFINNAEKGAVELDNPFILLVDKKVSNIRELLPTLEA 240

Qy        240 VAKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGE 299
              ||| ::|||||||:|||||||||||| :|| ::| |||||||:|||||:|||:||||
Db        241 VAKASKPLLIIAEDLEGEALATLVVNNMRGIVKVSAVKAPGFGDRRKAMLQDIAVLTGGT 300

Qy        300 AIFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSD 359
              | |: |::||   ::  |||||||| |::|||:||||: | |   || ||:||||| :||
Db        301 VISEEIGLELEKATVEDLGTAKRVVITKDDTTIIDGAGEEEGITGRVSQIKAQIEEATSD 360

Qy        360 YDREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAF 419
              ||:||||||:||| |||||| |||||| |||||||||||||:|||||||:||||| |
Db        361 YDKEKLQERMAKLAGGVAVIKVGAATEIEMKEKKDRVEDALHATRAAVEEGVVPGGGVAL 420

Qy        420 VRSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFN 479
              ||:   | ::   ::|:  |: :  |::| ||||| ||| | |:|:  |:  :  :|:|
Db        421 VRAASKLVELLGDNEDQNHGVKVALRAMEAPLRQIVTNAGDEASVVINAVKGGEGNYGYN 480

Qy        480 AASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMG 539
              ||:||| |:|: |::|| |||| ||| | |:| |::|||  :|| |: :       ||||
Db        481 AATGEYNDMIEMGILDPTKVTRSALQFAGSIAGLMITTEAMVAEIPKEEAAGAPDMGGMG 540

Qy        540 GMGGMDGM 547
              ||||| ||
Db        541 GMGGMGGM 548


RESULT 7
Q8CX48
ID   Q8CX48      PRELIMINARY;      PRT;   545 AA.
AC   Q8CX48;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Chaperonin groEL.
GN   GROEL OR SO0704.
OS   Shewanella oneidensis.
```

```
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Shewanella.
OX   NCBI_TaxID=70863;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MR-1;
RX   MEDLINE=22297686; PubMed=12368813;
RA   Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA   Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA   Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA   Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA   Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA   Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA   Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT   "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT   Shewanella oneidensis.";
RL   Nat. Biotechnol. 20:1118-1123(2002).
DR   EMBL; AE015516; AAN53782.1; -.
DR   TIGR; SO0704; -.
KW   Complete proteome.
SQ   SEQUENCE   545 AA;  57079 MW;  7591870B017DD521 CRC64;

  Query Match           65.7%;  Score 1780.5;  DB 16;  Length 545;
  Best Local Similarity  65.4%;  Pred. No. 2.5e-79;
  Matches  358;  Conservative   78;  Mismatches  108;  Indels    3;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||::|   || |:  ||: ||||||||||||||||::||||||:||||||||||||
Db          1 MAAKEVVFGNDARVKMLAGVNILANAVKVTLGPKGRNVVLDKSFGSPLITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| | :| ||||||||||||||||  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIVTEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||:|   || ::::   | | |||||||||||| :|| ||| || |||| |||||||
Db        121 DKAVIAAVAELKNLSQDCADSKAIAQVGTISANSDESIGQIIATAMEKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              : ||  |||||||:|||||||||:  ||     |||:|::|  :|||:::::|||||  :
Db        181 QALENELDVVEGMQFDRGYLSPYFINKPETGSVELDHPFVLLVDKKISNIRELLPILEGL 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :||||:|||||||||||||||| :|| ::| ||||||||:||||||:|:||||||
Db        241 AKTGKPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDVAILTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  |||||||||  |:||||:|| |:   |:||| ||: ||||::|||
Db        301 IAEEIGLELEKATLEDLGTAKRVVITKDNTTIIDGNGEQAQIEARVSQIKQQIEESTSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:||||||:||| |||||| |||||| |||||||:|||||||||:||||| | :
Db        361 DKEKLQERMAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    : |:: |::|:  |: |  |::| |||||| ||| | |:|    |:     :|:||
Db        421 RVASKIADVEVANEDQKHGVVIALRAMEAPLRQIATNAGEEASVVANNVKNGSGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              :  | |::: |::|| |||| ||| |||:| |::||| :||   || | |   |||||
Db        481 GNDTYGDMLEMGILDPTKVTRSALQFAASIAGLMITTEAMVAEL--PKADAP-DMGGMGG 537

Qy        541 MGGMDGM 547
              |||| ||
Db        538 MGGMGGM 544


RESULT 8
Q8GBB4
ID   Q8GBB4      PRELIMINARY;      PRT;   546 AA.
AC   Q8GBB4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL.
GN   GROEL.
OS   Azotobacter vinelandii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Azotobacter.
OX   NCBI_TaxID=354;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UWR;
RX   MEDLINE=93323980; PubMed=7687328;
RA   Ferat J.L., Michel F.;
RT   "Group II self-splicing introns in bacteria.";
RL   Nature 364:358-361(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UWR;
RA   Ferat J.-L., Legouar M., Michel F.;
RT   "An active group II intron has invaded the genus Azotobacter and is
RT   inserted within the essential groEL gene.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY057439; AAL25964.1; -.
SQ   SEQUENCE   546 AA;  56875 MW;  8E1F4CC93E42A619 CRC64;

  Query Match           65.7%;  Score 1780.5;  DB 2;  Length 546;
  Best Local Similarity  65.9%;  Pred. No. 2.6e-79;
  Matches  359;  Conservative   75;  Mismatches  110;  Indels    1;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:||: |   ||:|:  ||: ||:||| ||||||||||::||||:| ||||||||||||
Db          1 MAAKEVKFGDSARKKMLVGVNVLADAVKATLGPKGRNVVLDKSFGAPTITKDGVSVAKEI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              ||:|||||||:||:|| | :| ||||||||||||||  ||:| |||| ||| :||||
Db         61 ELKDKFENMGAQLVKDVASKANDEAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||| :|:  ||  : ||  | | |||||||||||| :|||||||| |||| |||||||
Db        121 DKATIAIVAELKSLAKPCSDSKAIAQVGTISANSDESIGNIIAEAMNKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              |||  | |||||:|||||||||:  |: || ||||| :|  :|||:::::||:|| |
Db        181 SGLENELSVVEGMQFDRGYLSPYFINKPDTMVAELDNPLLLLVDKKISNIRELLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  |||||:|||||||||||||| :|| ::| |||||||  :||||||: ||||||
Db        241 AKSGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFNDRRKAMLQVIAILTGATV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |  |: ||: :|  ||  | :|::|||||: |||   ||:| | ||| |||||||||
Db        301 ISEKVGLSLESATLEHLGKPKALVLNKENTTIMHGAGAQADIEAAVAQIRKQIEETSSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||||| |||||| |||||| ||| ||:|||||||||:|||||  | |
Db        361 DREKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEAALHATRAAVEEGVVPGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |::: :: :|  ::|:  |: ::||::| ||||| |||| | |:||:||::    |||||
Db        421 RALQAIEGLKGDNEDQNVGIALLRRAVEAPLRQIVANAGDEPSVVVDKVKQGSGNFGFNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMG 539
              ||| | |:|: |::|| |||| ||| |:|:  |::|||  :|:  | |    {{ {{{{
Db        481 ASGVYGDMIEMGILDPAKVTRSALQAASSIGGLMITTEAMVADIVEDKAAPAMPDMGGMG 540

Qy        540 GMGGM 544
              |||||
Db        541 GMGGM 545


RESULT 9
Q9RC20
ID   Q9RC20      PRELIMINARY;      PRT;   539 AA.
```

```
AC   Q9RC20;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Protein Cpn60 (60 kDa chaperonin) (groEL protein).
GN   CPN60.
OS   Bacillus sp. MS.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=96470;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS;
RA   Kondo A., Yohda M.;
RT   "Cpn60/10 from Bacillus strain MS.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AB028452; BAA88110.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   539 AA;  57346 MW;  95626C980E30C46F CRC64;

  Query Match           65.3%;  Score 1768;  DB 2;  Length 539;
  Best Local Similarity  65.0%;  Pred. No. 1e-78;
  Matches  354;  Conservative   85;  Mismatches   98;  Indels    8;  Gaps    3;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|||  |   :||   : |||||||:|||||||||||||||:|| ||||:|| |||::||||||
Db          2 AKEIKFSEEARRAMLRGVDKLADAVKVTLGPKGRNVVLEKKFGSPLITNDGVTIAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              || ||||||::| ||| ||:|:|||||||||||||: |||:| | || ||| |::||:|
Db         62 EDPFENMGAKLVAEVASKTNDVAGDGTTTATVLAQAMIREGLKNVTAGANPMGIRKGIEK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||     :||  |:|| : :: ||||   ||| :|  :|  :|||||  :||  ||||:||:||
Db        122 AVAVAVEELKAISKPIKGKESIAQVAAISA-ADEEVGQLIAEAMERVGNDGVITLEESKG 180

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              | |||||||:|||||:||| :|: |||    :|||||  :|||::::|:|||||| :
Db        181 FTTELDVVEGMQFDRGYVSPYMITDTEKMEAVLENPYILITDKKISNIQDILPILEQVVQ 240

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :||||||||:|||||||||||||||    |||||||||:||||||||||||||||||| |
Db        241 QGKPLLIIAEDIEGEALATLVVNKLRGTFTAVAVKAPGFGDRRKAMLEDIAILTGGEVIS 300

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: | :|:: :::||| | :||: |||||||||| || |||| ||||:|||:|:::||
Db        301 EELGRELKSATIASLGRASKVVVTKENTTIVDGAGDSERIKARINQIRAQLEETTSEFDR 360

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| ||| || |||||||:||:| |:|||||:|||||||| |||||| :
Db        361 EKLQERLAKLAGGVRVIKVGAATETELKERKLRIEDALNSTRAAVEEGIVAGGGTALMNV 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              :  |: |: ||  |: |: |::|||:|||| ||| |||::|||::  | | |||||:
Db        421 YNKVAAIE-AEGDEATGVKIVLRAIEEPVRQIAQNAGLEGSVIVERLKTEKPGIGFNAAT 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542
              ||: |:|:||::|| |||| |||||||||:: |||| :|:|||  |      || ||
Db        480 GEWVDMIEAGIVDPTKVTRSALQNAASVAAMFLTTEAVVADKPEENK------GGNAGMP 533

Qy        543 GMDGM 547
```

```
                | ||
Db        534 DMGGM 538


RESULT 10
O31198
ID   O31198      PRELIMINARY;      PRT;   546 AA.
AC   O31198;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Heat shock protein 58 (60 kDa chaperonin) (Protein Cpn60) (groEL
DE   protein).
GN   HSP58.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HY-1;
RA   Kim M.J., Ahn B.Y.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF032910; AAB86965.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   546 AA;  57969 MW;  15BC9EC0A8E2A6E3 CRC64;

  Query Match           65.3%;  Score 1768;  DB 2;  Length 546;
  Best Local Similarity  64.4%;  Pred. No. 1e-78;
  Matches  352;  Conservative   86;  Mismatches  101;  Indels    8;  Gaps    3;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :|:| ::  || ||  ||:|||||||||||||||||:| ||:| |||||||:|||||||
Db          2 AKDIEYNETARRKLLEGVNKLANAVKVTLGPKGRNVVIDKKFGAPTITKDGVTVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              ||  ||||||||||: ||:|:|||||||||:|||:|   ||:|  | || ||| |::||||||
Db         62 EDPLENMGAQMVKEVSTKTNDVAGDGTTTATILAQSIINEGLKNVTAGANPMSLKRGIDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              ||  |  : :         ::|:|| | :||||:| ||||:||:|| |||| |||||||||
Db        122 AVTAAVESIQKRAVKIENKKDIANVASISANNDNTIGNLIADAMDKVGKDGVITVEEAKS 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :||||||||:|||||:||| ||: | ||  |::|:||  :|||:||||:: |||:||:
Db        182 IETTLDVVEGMQFDRGYISPYMVTDAESMVATLNDPFILIYDKKISSMKDLIHILEKVAQ 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :||:||:|:|||||||:||| ||   :  |||||||||:|||:|||||||||||: |
Db        242 AGKPLVIISEEVEGEALATIVVNTLRKTISCVAVKAPGFGDRRKSMLEDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              || |:|||| :|  || | :| :|||||||::| |::::|: |: ||: |||:|:|:|||
Db        302 EDLGMKLENTTLQMLGRANKVTVDKENTTIIEGKGQTKEIQGRIGQIKKQIEDTTSEYDR 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATEVEMKEKKARVEDALSATRAAVEEGIVPGGGLTLLKA 421
              |||||||||| |||||||||||||| |||||| |||||:|||||||||||||||   :::
Db        362 EKLQERLAKLAGGVAVIHVGAATEVEMKEKKARVEDALSATRAAVEEGIVPGGGLTLLKA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              : :  :|  | ||  |  || |:||||:| | :||| |||::||  :  |  |||| :
Db        422 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 481
```

```
Db        422 QEAVGSLK-LDGDEATGAKIIFRALEEPIRMITSNAGLEGSVIVEHAKAKKGNEGFNALT 480

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMP-----GGG 537

Db        481 MVWEDMIQAGVVDPAKVVRSALQNAASIGSMILTTEVTITDKPD--KDAPNPMAGMGGGG 538

Qy        538 MGGMGGM 544

Db        539 MGGMGGM 545


RESULT 11
Q9EZV4
ID   Q9EZV4      PRELIMINARY;      PRT;   539 AA.
AC   Q9EZV4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN   GROEL.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC31783;
RA   Snapyan M., Gochikyan A., Weigel P., Sakanyan V.;
RT   "Effect of thermostable chaperonins on synthesis of proteins in
RT   vitro.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF274871; AAG44815.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   539 AA;  57293 MW;  2F51B6FFBB887428 CRC64;

  Query Match           64.8%;  Score 1754;  DB 2;  Length 539;
  Best Local Similarity   64.6%;  Pred. No. 4.9e-78;
  Matches  352;  Conservative   87;  Mismatches   98;  Indels    8;  Gaps    3;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62

Db          2 AKEIKFSEEARRAMLRGVDKLADAVKVTLGPKGRNVVLEKKFGSPLITNDGVTIAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122

Db         62 EDPFENMGAKLVVEVASKTNDVAGDGTTTATVLAQAMIREGLKNVTAGANPMGIRKGIEK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182

Db        122 AVAVAVEELKAISKPIQGKESIAQVAAISA-ADEEVGQLIAEAMERVGNDGVITLEESKG 180

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242

Db        181 FTTELDVVEGMQFDRGYASPYMITDTEKMEAVLENPYILITDKKISNIQDLLPILEQVVQ 240

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302

Db        241 QGKPLLIIAEDVEGEALATLVVNKLRGTFTAVAVKAPGFGDRRKAMLEDIAILTGGEVIS 300

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362

Db        301 EELGRELKSTTIASLGRASKVVVTKENTTIVEGAGDSDRIKARINQIRAQLEETTSEFDR 360

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422

Db        361 EKLQERLAKLAGGVAVIKVGAATETELKERKLRIEDALDSTRAAVEEGIVAGGGTALMNV 420

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482

Db        421 YSKVAAIE-AEGDEATGVKIVLRAIEEPVRQIAQNAGLEGSVIVERLKTEKPGIGFNAAT 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGGMG 542

Db        480 GEWVNMIEAGIVDPTKVTRSALQNAASVAAMFLTTEAVVADKPEENK------GGNNGMP 533

Qy        543 GMDGM 547

Db        534 DMGGM 538


RESULT 12
Q93FU8
ID   Q93FU8      PRELIMINARY;      PRT;   545 AA.
AC   Q93FU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL-like protein (60 kDa chaperonin) (Protein Cpn60) (groEL
DE   protein).
GN   GROEL.
OS   Enterobacter aerogenes (Aerobacter aerogenes).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Enterobacter.
OX   NCBI_TaxID=548;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yoshida N., Oeda K., Watanabe E., Mikami T., Fukita Y., Nishimura K.,
RA   Komai K., Matsuda K.;
RT   "A GroEL homolog produced by endosymbionts in antlions as a paralytic
RT   toxin against insects.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF306521; AAL09389.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   545 AA;  57023 MW;  DB3B7A017DEDA85F CRC64;

  Query Match           64.3%;  Score 1742;  DB 2;  Length 545;
  Best Local Similarity   64.7%;  Pred. No. 1.9e-77;
  Matches  353;  Conservative   73;  Mismatches  116;  Indels    4;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60

Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120

Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180

Db        121 DKAVVAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
```

```
                ||:  |||||||:||||||||||:  ||    ||::|:||  :|||:::::|||:|| |
Db        181 TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||||||||| :|| ::| ||||||||:||||||:||| ||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  || ||||||:|: |||:|| |    |: || ||| |||| :|||
Db        301 ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGDEAAIQGRVTQIRQQIEEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |||||||||:||| |||||| |||||| |||||| ||||||:|||||||||:| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    :  :|  ::|:  |: :  |::| |||||  | | | |:|   |:      :|:||
Db        421 RVASKIAGLKGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKAGDGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG--GGM 538
              |: || ::|  |::|| |||| ||| ||||| |::|||| : :   || | |  | |||
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDL--PKSDAPDLGAAGGM 538

Qy        539 GGMGGM 544
              ||||||
Db        539 GGMGGM 544


RESULT 13
O33688
ID   O33688      PRELIMINARY;      PRT;   538 AA.
AC   O33688;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60) (Fragment).
GN   GROEL.
OS   primary endosymbiont of Sitophilus oryzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae.
OX   NCBI_TaxID=2342;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SFr;
RX   MEDLINE=98042468; PubMed=9367844;
RA   Charles H., Heddi A., Guillaud J., Nardon C., Nardon P.;
RT   "A molecular aspect of symbiotic interactions between the weevil
RT   Sitophilus oryzae and its endosymbiotic bacteria: over-expression of a
RT   chaperonin.";
RL   Biochem. Biophys. Res. Commun. 239:769-774(1997).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF005236; AAB97670.1; -.
DR   HSSP; P06139; 1JON.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
FT   NON_TER     538     538
SQ   SEQUENCE   538 AA;  56683 MW;  EFD17668C0F09D15 CRC64;

  Query Match             64.1%;  Score 1736;  DB 2;  Length 538;
  Best Local Similarity   64.0%;  Pred. No. 3.7e-77;
  Matches  348;  Conservative   81;  Mismatches  109;  Indels    6;  Gaps    2;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |   || |: ||| ||:|||||||||||||::||||:|||||||||||:||
Db          1 MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPVITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||| | :| ||||||||||||:|  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQSIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||:|  :||  ::  |  | |||||||||||| |:| :|||||||| |||||||
Db        121 DKAVIAAVEELKKLSVPCSDSKAIAQVGTISANSDETVGTLIAEAMAKVGKEGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||:  |||||||:|||||||||||  ||     ||::|:||  :|||:::::|||:|| |
Db        181 SGLQDELDVVEGMQFDRGYLSPYFVNKPETGAIELESPFILLADKKISNIREMLPVLEAV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :|||||||||||||||||||| :|| ::| ||||||||:||||||:||| ||||
Db        241 AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIAILTAGTV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |: |::||  :|  :| |||||| |: |||:|| |    | :|| ||  | :| :|||
Db        301 ISEEIGLELEKATLEDMGQAKRVVITKDTTTIIDGVGDKALIDSRVTQINQQRDEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              ||||||||:||| |||||| |||||| |||||| ||||||:|||||||||:| ||| | :
Db        361 DREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    : :::  ::|:  |: : ||::| ||||| |||| | |::  ||:  :   |:||
Db        421 RVANSIAELRGDNEDQNVGIKVARRAMEAPLRQIVANAGEEPSVIANKVKAGEGNTGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPGGGMGG 540
              |: || ::|  |::|| |||| ||| |||:| |::|||| : :  :||:| |     :||
Db        481 ATEEYGNMIDMGILDPTKVTRSALQYAASIAGLMITTECMVTD--QPKEDKP----DLGG 534

Qy        541 MGGM 544
              ||||
Db        535 MGGM 538


RESULT 14
Q8RNU2
ID   Q8RNU2      PRELIMINARY;      PRT;   546 AA.
AC   Q8RNU2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Heat shock protein B subunit (60 kDa chaperonin) (Protein Cpn60)
DE   (groEL protein).
GN   HSPB.
OS   Helicobacter pylori.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CH-CTX1;
RX   MEDLINE=20123532; PubMed=10660136;
RA   Opazo P., Muller I., Rollan A., Valenzuela P., Yudelevich A.,
RA   Garcia-de la Guarda R., Urra S., Venegas A.;
RT   "Serological response to Helicobacter pylori recombinant antigens in
RT   Chilean infected patients with duodenal ulcer, non-ulcer dyspepsia and
RT   gastric cancer.";
RL   APMIS 107:1069-1078(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CH-CTX1;
RA   Muller I., Medina-Selby A., Palacios J.L., Martinez P., Opazo P.,
RA   Bruce E., Mancilla M., Valenzuela P., Yudelevich A., Venegas A.;
RT   "Cloning and comparison of ten gene sequences of a Chilean
RT   Helicobacter pylori strain with other Helicobacter pylori strains
RT   revealed higher amino acid sequence variability for VacA and CagA
RT   virulence factors.";
```

```
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AF479030; AAL86900.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   546 AA;  58242 MW;  B3DAE13619F26B32 CRC64;

  Query Match             63.6%;  Score 1722.5;  DB 2;  Length 546;
  Best Local Similarity   64.1%;  Pred. No. 1.7e-76;
  Matches  350;  Conservative   73;  Mismatches  120;  Indels    3;  Gaps    2;

Qy          3 SKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEIEL 62
              :||| |   ||  |  || :| :|||||:||:||||:|:||:|:| ||||||||||||||
Db          2 AKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVAKEIEL 61

Qy         63 EDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGIDK 122
              |||||:|||||  ||:| |||||||||||| :|::||:: : || ||: :|||:|| 
Db         62 SCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPIEVKRGMDK 121

Qy        123 AVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEAKG 182
              | |:  ||   :|    ::|| || |||||||  || :||:|| |||| ||||||||||
Db        122 AAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGKDGVITVEEAKG 181

Qy        183 LETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQVAK 242
              :| |||||||:|||||||||||||| |||  :||| |||  :|||:||||:||:||:  |
Db        182 IEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKISSMKDILPLLEKTMK 241

Qy        243 VNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEAIF 302
              :|||||||:||||| |||||||| | : ||||||||:||| ||:|||||||: |
Db        242 EGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRKEMLKDIAILTGGQVIS 301

Qy        303 EDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDYDR 362
              |: |: |||  :  || | |:||||:||||||| | |:|:| || ||: ||  |:||||:
Db        302 EELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSDDVKDRVAQIKTQIASTTSDYDK 361

Qy        363 EKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFVRS 422
              |||||||||| |||||| |||:| |||||||||:|||:||:|||||||| ||| | :|:
Db        362 EKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDALSATKAAVEEGIVIGGGAALIRA 421

Qy        423 IKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNAAS 482
              : :       ||| |  || |::: || ||| |||||:| :|| :| : :  |||||::
Db        422 AQKVH--LNLHDDEKVGYEIIMRAIKAPLAQIAINAGYDGGVVVNEVEKHEGHFGFNASN 479

Qy        483 GEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKDMPMPG-GGMGGM 541
              |:| |: | |:||| || ||||||| ||:|||||  : |  | |     ||  ||||||
Db        480 GKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTTEATVHEIKEEKAAPAMPDMGGMGGM 539

Qy        542 GGMDGM 547
              ||| ||
Db        540 GGMGGM 545


RESULT 15
Q8KIW5
ID   Q8KIW5       PRELIMINARY;      PRT;   548 AA.
AC   Q8KIW5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   groEL protein (60 kDa chaperonin) (Protein Cpn60).
GN   GROEL.
OS   Buchnera sp. 150.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=168690;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=150 Neuville-1;
RA   Delmotte F., Sabater B., Simon J.C., Latorre A.;
RT   "Coevolution between mitochondrialand simbiotic DNA at intraspecific
RT   level.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=150 Neuville-1;
RA   Delmotte F.;
RL   Thesis (2001), University of Valencia, Spain.
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   EMBL; AJ344978; CAC86959.1; -.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   ATP-binding; Chaperone.
SQ   SEQUENCE   548 AA;  57853 MW;  659886423C19529D CRC64;

  Query Match             63.3%;  Score 1714.5;  DB 2;  Length 548;
  Best Local Similarity   61.8%;  Pred. No. 4.3e-76;
  Matches  338;  Conservative   92;  Mismatches  114;  Indels    3;  Gaps    1;

Qy          1 MASKEILFDAKAREKLSRGVDKLANAVKVTLGPKGRNVVIEKSFGSPVITKDGVSVAKEI 60
              ||:|:: |  :|| |: |||: ||:|||||||||||||::||||:| |||||||||:||
Db          1 MAAKDVKFGNEARIKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPSITKDGVSVAREI 60

Qy         61 ELEDKFENMGAQMVKEVAPKTSDIAGDGTTTATVLAQAIYREGVKLVAAGRNPMAIKRGI 120
              |||||||||||||||||| | :| |||||||||:|||:|  ||:| |||| ||| :||||
Db         61 ELEDKFENMGAQMVKEVASKANDAAGDGTTTATLLAQSIVNEGLKAVAAGMNPMDLKRGI 120

Qy        121 DKAVVAVTKELSDITKPTRDQKEIAQVGTISANSDTTIGNIIAEAMAKVGKGGVITVEEA 180
              ||||::  :|| ::: |  | | | ||||||||:|  :| :|||| |||  |||||||
Db        121 DKAVISAVEELKNLSVPCSDSKAITQVGTISANADEKVGALIAEAMEKVGNDGVITVEEG 180

Qy        181 KGLETTLDVVEGMKFDRGYLSPYFVTNPEKMVCELDNPYILCNEKKITSMKDMLPILEQV 240
              ||:  |:||:||:||||||||||:  ||  | ||:|||||  :|||:::::|||||| |
Db        181 TGLQNELEVVKGMQFDRGYLSPYFINKPETGVVELENPYILMADKKISNVREMLPILESV 240

Qy        241 AKVNRPLLIIAEDVEGEALATLVVNKLRGALQVVAVKAPGFGERRKAMLEDIAILTGGEA 300
              ||  :||||:||:|||||||||||| :|| ::| |||||||:||||||:||::|||| 
Db        241 AKSGKPLLIISEDLEGEALATLVVNSMRGIVKVAAVKAPGFGDRRKAMLQDISVLTGGSV 300

Qy        301 IFEDRGIKLENVSLSSLGTAKRVVIDKENTTIVDGAGKSEDIKARVKQIRAQIEETSSDY 360
              | |:  : ||  :|  || |||||:|: |||: | |: : |::|: ||| :|:| :|||
Db        301 ISEELAMDLEKSTLEDLGQAKRVVINKDTTTIIGGVGEKQAIQSRISQIRQEIQEATSDY 360

Qy        361 DREKLQERLAKLVGGVAVIHVGAATETEMKEKKDRVEDALNATRAAVEEGIVPGGGTAFV 420
              |:||| |||||| |||||: |||||| |||||| |||||||:|||||||||:| ||| | |
Db        361 DKEKLNERLAKLSGGVAVLKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALV 420

Qy        421 RSIKVLDDIKPADDDELAGLNIIRRSLEEPLRQIAANAGYEGSIVVEKVREPKDGFGFNA 480
              |    : :::  ::|:  |: :  |::| ||||| :|:| | |:|   |:: |  :|:||
Db        421 RVAGKISNLRGHNEDQNVGIRVALRAMEAPLRQIVSNSGEEPSVVTNNVKDGKGNYGYNA 480

Qy        481 ASGEYEDLIKAGVIDPKKVTRIALQNAASVASLLLTTECAIAEKPEPKKD---MPMPGGG 537
              |: || |:|  |::|| |||| ||| ||||| |::|||| : : |:  |       | ||
Db        481 ATDEYGDMIDFGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKEDKTSDASSSPAGG 540

Qy        538 MGGMGGM 544
              |||||||
Db        541 MGGMGGM 547
```

Search completed: January 28, 2004, 13:17:17
Job time : 42 secs